

Open Research Online

The Open University's repository of research publications
and other research outputs

Epidemiological and molecular studies on chronic HBV infection in Gambian families

Thesis

How to cite:

Dumpis, Uga (2001). Epidemiological and molecular studies on chronic HBV infection in Gambian families. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2001 Uga Dumpis

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.0000f98d>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Epidemiological and molecular studies on chronic HBV infection in Gambian families

Uga Dumpis, Dr.Med.

A thesis submitted in partial fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

August 2000

Institute of Molecular Medicine, University of Oxford with the MRC laboratories, Fajara, The Gambia


DATE OF SUBMISSION : 15 AUGUST 2000
DATE OF AWARD : 22 JANUARY 2001

ProQuest Number: C808699

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest C808699

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Summary

Over 600 chronic hepatitis B virus (HBV) carriers were identified in The Gambia. 142 families with multiple affected siblings were recruited for this study. The prevalence of HBsAg carriage in siblings of the carriers was significantly higher than in normal population with the same exposure rate to the HBV infection, which could be considered as the sign of the family clustering of the disease. Significant correlation between birth order of a sibling and the risk of becoming a carrier was found. Relatively low prevalence in mothers of HBV carriers in these families was suggestive that mother to child HBV transmission did not play a major role. The index cases have been assumed to come from within a family unit or from sources outside an immediate family, such as other families or inhabitants of the same compound or village. I have attempted to define these routes of HBV transmission by phylogenetic analysis of sequences from the entire HBV pre-core/core and surface gene regions. Pre-core/core gene sequences were found to be more informative for this purpose. Despite the overall conserved nature of the pre-core/core sequences, distinct clusters with high bootstrap support were easily defined at the family and village level, but not on the wider geographical distribution level. Over one third of the families studied had a strong evidence of intrafamilial HBV transmission. Therefore clustering of HBV infection in the Gambian families was confirmed with epidemiological data and the hypothesis on horizontal HBV transmission between West African children for the first time was supported by sequencing data and phylogenetic analysis. A large number of unique Gambian HBV core promoter/pre-core/ core and surface gene sequences were obtained and their molecular characteristics described. The analysis also revealed for the first time that HBV genotypes A and E circulate in The Gambia, with the clear predominance of the latter.

Dedication

I would like to dedicate this work to all Gambian people who participated in this study.

Special thanks to my parents and my beautiful girlfriend Vita for their patience and love during long years of my absence.

Acknowledgements

This work would not have been possible without the feedback from different experienced scientists, friends and unbelievably co-operative Gambian people.

I wish to acknowledge with my sincere thanks the help and direction of my supervisors Prof. Adrian Hill and Prof. Andy Hall during the progress of this work.

I would also like to thank Dr. Hilton Whittle for his encouragement and inspiring assistance to change different scientific approaches in order to test the hypothesis. His knowledge and experience of work in the Gambian environment was crucial for the success for this study.

A special thank you to Dr. Michael Wansbrough, Mrs Maimuna Mendy, Mrs Adam Yeng, Mr Lamin Giana, Mr Joseph Bass and Mr Lamin Bojang for assisting with the field and the laboratory work. Without their help, sample collection would not be possible.

My great thanks to Dr. Edward Holmes for helping to find the best statistical approach for the analysis of phylogenetic trees.

Patience and politeness of Dr. Peter Karayiannis in teaching me basics of molecular techniques were above all expectations and without his supervision, success of the molecular work would not be possible.

Finally, I want to thank all my teachers in Latvian Medical Academy and Soros Scholarship Foundation which provided me with wonderful opportunity to go to Oxford University and do research work for one year, which was the starting point for my further development as a scientist.

List of contents

1. Summary	15
2. Introduction	17
2.1 General background of the study	17
2.2 Hepatitis B virus	18
2.3 Transmission of hepatitis B virus.....	33
2.4 Natural history of HBV infection	38
2.5 HBV and hepatocellular carcinoma	49
2.6 Chronic HBV carriage	51
2.7 Molecular epidemiology of the intrafamilial transmission of HBV infection	69
2.8 Phylogenetic methods	71
2.9 Geographical setting of the study	76
3. Methods	
3.1 Field work	79
3.2 Laboratory work	81
4. Human epidemiology of HBV and characteristics of the chronic carriers	94
4.1 Introduction	94
4.2 Results	95
4.3 Discussion	112
5. Molecular epidemiology of HBV in the Gambia	116
5.1 Introduction	116
5.2 Optimisation of OCR for the surface and pre-core/core regions of HBV.....	119
5.3 Validation of PCR results.....	122
5.4 Sequences.....	125

6. Phylogenetic tree analysis of HBV sequences in families	144
6.1 Introduction	144
6.2 Phylogenetic analysis of surface gene fragments.....	145
6.3 Phylogenetic analysis of pre-core/core gene fragments.....	148
6.4 Detailed analysis on family and village clustering.....	151
6.5 Phylogenetic analysis according to HBeAg status.....	157
6.6 Sequences from mothers and their children	162
6.7 Sequences from the village Manduar	164
6.8 Statistical analysis of the clustering in families and villages	166
6.9 Discussion	170
7. General Discussion	177
8. Future plans	194
9. Appendices	196
10. References	205

List of figures

1. Structure of the α determinant of the HBsAg .
2. Serological findings during the typical course of an acute HBV infection.
3. Map of The Gambia
4. ALT levels in HBsAg positive and HBsAg negative patients.
5. AST levels in HBsAg positive and HBsAg negative patients.
6. Open reading frames of HBV and positioning of primers used in this study.
7. Neighbour- joining phylogenetic tree on pair sequence sets from the same samples.
8. 1% agarose gel with PCR products obtained with primers S3-S4 and stained with ethidium bromide.
9. 1% agarose gel with PCR products obtained with primers M3-Pol8 and M3-5C and stained with ethidium bromide.
10. 1% agarose gel with PCR products obtained with primers M3-5C, M3-8C, C5-C6 and stained with ethidium bromide.
11. Comparison of the Gambian consensus sequences from the α determinant of genotypes A and E to sequences of other genotypes.
12. Neighbour joining phylogenetic tree representing sequences from precore/core regions from all known human genotypes and two Gambian consensus sequences from each genotype.
13. Neighbour joining phylogenetic tree representing sequences from the S gene (nt 509-810) from all subtypes, genotypes and the Gambian consensus sequences from each genotype.
14. Geographical distribution of HBV genotypes E and A in The Gambia.

15. Amino acid sequences from the region of the S gene obtained from 36 patients.
16. Predicted secondary structure of the pregenomic HBV-RNA encapsidation signal showing the base pairing in genotypes A and E and nucleotide changes in precore sequences from The Gambia.
17. Distribution of amino acid changes in the Core gene and their relation to previously published B and T-cell epitopes
18. Pre-core/core sequences with amino acid substitutions according to their E antigen status.
19. Neighbour joining tree on 38 surface gene nucleotide sequences.
20. Rooted neighbour joining phylogenetic tree on 143 pre-core/core gene nucleotide sequences.
21. Maximum-likelihood tree with neighbour-joining bootstrap values on 99 genotype E precore, core gene sequences.
22. Maximum-likelihood tree with neighbour-joining bootstrap values on 14 genotype A precore/core gene sequences.
23. Maximum likelihood tree on HBeAg positive samples for pre-core/core sequences.
24. Maximum likelihood tree on HBeAg negative samples.
25. Maximum likelihood tree on sequences from families with a HBsAg positive mother and their children.
26. Maximum likelihood tree (with bootstrap node values) on pre-core/core sequences from the village Manduar.
27. Example of how the maximum likelihood tree on sequences from HBeAg positive was transformed to a model tree with combined family+village cluster.

List of tables

1. Geographical distribution of HBV subtypes versus genotypes.
2. Serological findings for the different stages of HBV infection.
3. Prevalence of HBsAg and HBeAg in two Gambian villages.
4. Family clustering of HBsAg carriage in Keneba and Manduar
5. Prevalence of HBeAg and HBV DNA positivity in chronic HBsAg carriers according to their age.
6. Interrelationships among the five best-known models for estimating the number of nucleotide substitutions among a pair of DNA sequences.
7. Buffers and solutions
8. DNA extraction mix
9. PCR reaction mix
10. Thermo-cycling parameters of PCR reaction
11. Known carriers bled in previous studies
12. Proportion of family members bled in families, where only one carrier was identified previously.
13. Prevalence of ethnic groups in study subjects in comparison with the overall prevalence in The Gambia.
14. Number of families with multiple affected siblings HBV chronic carriers.
15. Prevalence of HBsAg positivity in parents from all families with at least one sibling carrier and from the families with multiple siblings.
16. HBsAg status according to sex of the siblings of the chronic carrier.
17. Relation of birth order adjusted for family size and mother carrier status
18. Prevalence of positive HBeAg in HBsAg in carriers related to age and sex.

19. PCR positivity according to HBeAg status
20. Prevalence of PCR positivity according to age and sex.
21. Levels of ALT in HBsAg positive and HBsAg negative persons
22. ALT values and HBsAg and HBeAg status in relation to the gender of the patient.
23. Levels of AST in HBsAg positive and HBsAg negative persons
24. AST levels and HBsAg and HBeAg status in relation to the gender of the patient.
25. Relation of sex and clearance of HBsAg over 6-10 year time period
26. Relation of age and clearance of HBsAg over 6-10 year time period.
27. Relation between HBeAg markers and number of mutations in the Core gene
28. Amino acid substitutions in the core gene found in this study and in relation to previous reports.
29. Relation of amino acid substitutions found in the Gambian samples to normal sequences of other genotypes or subtypes.
30. Summary of data and substitution parameters of the maximum likelihood phylogenetic trees constructed.
31. Log likelihoods for three different “model” phylogenetic relationships of HBV sequences (family, village and village+family) and a null distribution constructed using random tree topologies.

List of appendices

1. Consent form.
2. Family tree field questionnaire.
3. Field questionnaire.
4. Result form for parents.
5. Main serological markers and age of the family members of the previously known HBV chronic carriers.
6. Example of the FASTA sequence alignment format.
7. Example of the NEXUS sequence alignment format.
8. Example of the log likelihood values for different simulated trees.

Abbreviations

aa	amino acid
AFB1	aflatoxin B1
ALT	alanine-aminotransferase
anti-HBs	antibody to hepatitis B core antigen
anti-HBe	antibody to hepatitis B 'e' antigen
anti-HBs	antibody to hepatitis B surface antigen
AST	aspartate-aminotransferase
BCP	basal core promoter
CTL	cytotoxic lymphocyte
DNA	deoxiribonucleic acid
dNTPs-	equimolar mixture of dATP, dCTP, dGTP and dTTP
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytidine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
dTTP	thymine deoxyribonucleoside triphosphate
ELISA	enzyme linked immunosorbent assay
HBcAg	hepatitis B core antigen
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBxAg	hepatitis B X antigen
HCC	hepatocellular carcinoma
IFN	interferon
IG	immunoglobulin
IgG	immunoglobulin G
IgM	immunoglobulin M
IL-10	interleukin 10
IU	international unit
MBP	mannose binding protein

mM	milimolar
ml	mililiter
ML	maximum likelihood
mRNA	messenger RNA
NJ	neighbour joining
nmol	nanomolar
NO	nitric oxide
nt	nucleotide
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RIA	radioimmunoassay
RNA	ribonucleic acid
RIA	radioimmunoassey
RPA	reverse passive agglutination
RT	reverse transcriptase
SD	standart deviation
SDS	sodium dodycylsulfate
TNF	tumour necrosis factor
tRNA	transport RNA
Ts	transition
Tv	transversion
UK	United Kingdom
WHO	World Health Organization
WHV	woodchuck hepatitis virus

1 Summary

Over 600 chronic HBV carriers were identified during my work in The Gambia. 142 families with multiple affected siblings were recruited for this study. The prevalence of HBsAg carriage in siblings of the carrier was significantly higher than in a normal population with the same exposure rate to the infection, which could be considered as the sign of the family clustering of the disease. The transmission of hepatitis B virus (HBV) in West Africa was previously shown to occur during the first two years of life. Relatively low prevalence in mothers of the carriers in these families was suggestive that the mother to child transmission did not play a major role. The index cases in such events have been assumed to come from within the family unit or from sources outside the immediate family, such as other families or inhabitants of the same compound or village. I have attempted to define these routes of transmission by phylogenetic analysis of sequences from the entire pre-core/core region of the virus and sequences from the surface gene region. With surface gene sequences analysis did not seem feasible because of the highly conserved nature of the sequences. Precore/core gene sequences were found to be more informative. Despite the overall conserved nature of the sequences, distinct clusters with high bootstrap support were easily defined at the family and village level, but not on a wider geographical separation. Over one third of the families studied had a strong evidence of intrafamilial transmission.

Therefore the family clustering of HBV infection in Gambian families was confirmed with epidemiological data and the hypothesis on horizontal transmission of the virus in West Africa was for the first time supported by sequencing data and phylogenetic analysis. A large number of unique Gambian HBV S and C gene sequences were obtained and their

molecular characteristics described. The analysis also revealed for the first time that genotypes A and E circulate in The Gambia, with the latter predominating.

2 Introduction

2.1 General background of the study

Hepatitis B is one of the most common, serious infectious diseases in the world. More than 350 million people worldwide are chronic carriers (Lee, 1997). The World Health Organisation estimates that the infection leads to more than one million deaths every year. Each year approximately 300 000 people will become infected with the Hepatitis B virus (HBV). HBV is believed to cause as much as 80% of the global hepatocellular carcinoma (HCC) cases and to be second to tobacco as an environmental carcinogen.

Although there is an effective vaccine for the prevention of HBV, it is of no value to those already infected. Evidence for usage of the conventional HBV vaccine for treatment is inconclusive. While treatments are currently available for the chronic infection, these have considerable limitations in terms of treatment efficacy. The cost, patient compliance and administration of both the treatment and preventative therapy for HBV infection can make attainment of lower prevalence rates using present resources, an expensive and difficult project. The aim now must be to develop more cost effective vaccines and treatment that can be used in developing countries.

The work carried out for this thesis is part of a larger collaboration between Professor Adrian Hill's group in Oxford University, Professor Howard Thomas group in Imperial College School of Medicine at St.Mary's, London and Roche Discovery Ltd (Welwyn Garden City). The aims of this collaboration were to investigate the genetic determinants of susceptibility to persistent infection and disease with HBV. For these genetic family studies a cohort of sibling pairs who were persistently infected with

hepatitis B were recruited. The thesis focuses on the epidemiological, serological and molecular aspects of family clustering of HBV infection in The Gambia

2.2 Hepatitis B virus

2.2.1 Viral genome

HBV belongs to a group of hepatotropic DNA viruses (hepadnaviridae) that include the woodchuck, ground squirrel, Peking duck and heron hepatitis viruses.

The infectious Dane particle has a diameter of about 42 nm and is composed of a lipid bilayer envelope containing hepatitis B surface antigen (HBsAg) and an internal nucleocapsid. The nucleocapsid, which consists of the core protein, contains the DNA genome of the virus, which is between 3182 and 3221 bases long and partially double stranded with an associated DNA polymerase/reverse transcriptase (Tiollais et al, 1981). The numbering of the bases starts in most publications at the cleavage site for the restriction enzyme EcoRI or at homologous sites if a particular genome type does not have such an EcoRI site.

The HBV genome has four opening reading frames (ORFs), which are encoded by the same DNA strand. The genomes of hepadnaviruses seem to have evolved towards minimal length. The genome is not much longer than its longest ORF P, which encodes the viral DNA polymerase and its accessory functions. ORF S encodes the three HBsAg proteins. ORF C encodes the HBeAg protein and HBcAg protein, whilst ORF X encodes the X protein.

2.2.1.1 Viral proteins

2.2.1.1.1 Envelope proteins (HBsAg, Pre-S1, Pre-S2)

The envelope proteins (pre-S1, pre-S2 and S) are encoded by the PreS/S gene. The HBsAg proteins are synthesised and assembled at the membrane of the endoplasmic reticulum and bud into the lumen . From the ER the particles are transported to the cell membrane where they are released by vesicles. Excess HBsAg occurs abundantly in serum as small (22nm) spherical or filamentous, non-infectious particles. In natural infection, the ratio of non-infectious HBsAg particles to virions is about 1000 to 1. Three different HBsAg are synthesised: the large HBsAg (encoded by the pre-S1, pre-S2 and S genes), the middle HBsAg (encoded by the pre-S2 and S genes) and the major HBsAg (encoded by the S gene) (Tiollais et al, 1985). The S gene region, encompassing roughly codons 124 to 147 encodes the highly immunogenic “a” determinant. This epitope induces virus neutralising anti-HBs response after natural infection (Waters et al, 1992) or vaccination (Halliday et al, 1992).

2.2.1.1.2 HBcAg and HBeAg

The C ORF contains two in-frame initiation codons. Initiation of translation of the first one results in the production of the PreC/C protein from which the soluble HBeAg protein is derived, following proteolytic processing of its amino and carboxyl termini. The first 19 aa of the precore constitute a signal peptide, which is removed by signal peptidases at the ER membrane. The remainder of the protein is released into the lumen where it undergoes further processing of its carboxyl end and what remains is the HBcAg, which is released into the serum.

Initiation of translation at the second AUG results in protraction of the nucleocapsid protein of the virus of HBcAg.

2.2.1.1.3 Viral DNA polymerase/reverse transcriptase

ORF P overlaps with ORF C and ORF S and covers almost the entire length of the genome. This encodes the DNA polymerase of the viruses, which fulfils several functions: reverse transcription, priming of DNA synthesis, DNA polymerase activity, and RNase H activity.

2.2.1.1.4 HBxAg

The fourth ORF encodes the small protein X. The main function of X is thought to be as transcriptional activator and it has been shown to transactivate a number of cellular and viral promoters (Rossner et al, 1992). The X region overlaps with the Enhancer II/ core-promoter region, which directs the transcription of precore/core sequences and is essential for the production of pregenomic RNA.

2.2.1.2 Genotypes and subtypes

Seven genotypes A-G of HBV, which differ by more than 8% in protein homology, have been identified (Okamoto et al, 1988; Norder et al, 1994; Stuyver et al, 2000). A difference in protein sequence of 4% at the level of the S gene is considered to be sufficient to differentiate between genotypes (Norder et al, 1992).

At present the clinical value of genotyping and subtyping is uncertain, but there are indications suggesting that genotyping may be of clinical importance. A Japanese study showed that severe liver disease is more common in carriers with subtype *adr* (genotype C) than in those with *adw* (mainly genotype B in Japan) (Shiina et al, 1991). Swedish data associate C-1858 strains with more severe liver damage (Lindh et al, 1996). Another study showed that patients infected with genotype A responded better to interferon therapy

(Zhang, 1996). A Spanish study found that patients with genotype D are more likely to have persistent HBV infection by selection of pre-core mutants (Rodriguez-Frias et al, 1995). Genotypes B and C, which are dominant in high-prevalence areas in East Asia, may be associated with prolonged HBeAg carriage and vertical transmission. In contrast, horizontal transmission is more important in sub-Saharan Africa and the Mediterranean area where genotypes A, D and E are common. Finally, a recent study associated genotype C with more severe liver disease and genotype B with the development of HCC in young Taiwanese (Kao et al, 2000).

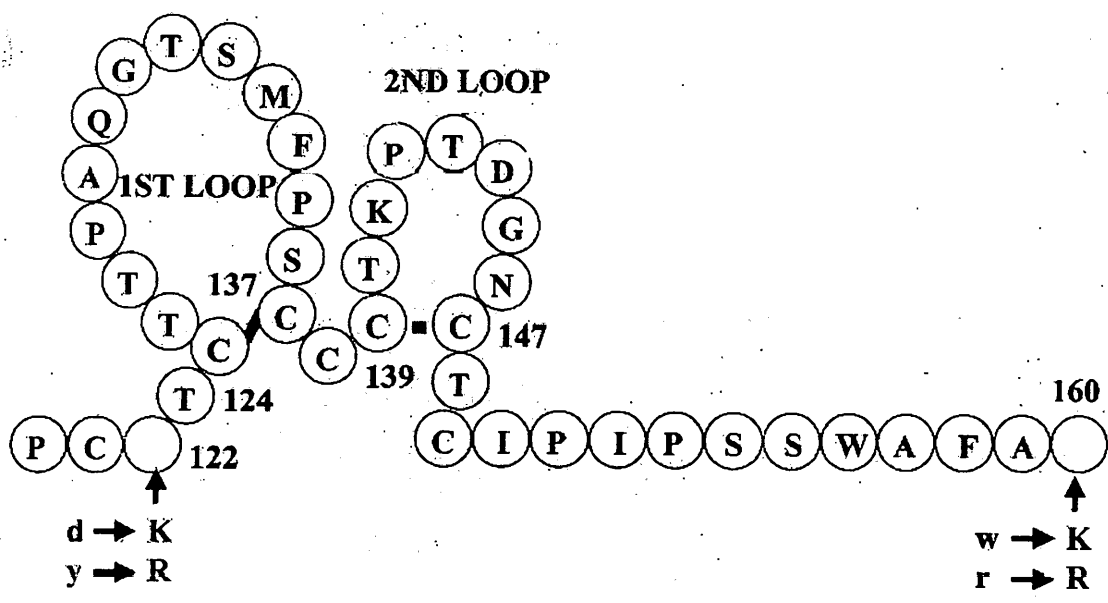


Figure 1. Structure of the α determinant of the HBsAg . (Karayiannis, personal communication)

HBsAg carries a group-specific determinant *a*, common to all subtypes of this antigen, and two additional subtypic determinants, *d* or *y* and *w* or *r*, depending on the amino acids in position 122 and 160 (See figure 1). As a result, four major subtypes of HBsAg are recognised: *adw* (K in positions 122 and 160), *adr* (K in 122, R in 160), *ayw* (R in 122, K in 160) and *ayr* (R in 122, R in 160). (Le Bouvier et al, 1971; Bancroft et al, 1972). With the description of 4 sub-determinants (Sandler et al, 1978) of *w* and with identification of the *q* determinant (Magnius et al, 1975) the number of subtypes increased to nine: *ayw1*, *ayw2*, *ayw3*, *ayw4*, *ayr*, *adw2*, *adw4*, *adrq-* and *adrq+*.

Little is known about the distribution of HBV genotypes and subtypes in West Africa. A vaccine escape mutation from Lys to Glu at position 141 has been reported from West Africa (Devi-Karthigesu et al, 1994). This mutant has only been found in association with *ayw4* subtype and genotype E. Other viral genotypes apart from E from West Africa have not been described (Lindh et al, 1997), but there are many reports about high prevalence of genotypes A (serotype *ayw1*) and D in other parts of Africa (Norder et al, 1993).

Sequencing of the *a* determinant region of the S gene is sufficient for determining viral genotype and subtype. Classification of HBV genomes on the basis of the sequence of the S gene alone is feasible, because such results are in agreement with the previous classification based on the complete genome (Norder, 1993).

Genotype	Subtype	Geographical distribution
A	adw2	Europe, USA
A	ayw1	Africa
B	adw2	Asia, Pacific
B	ayw1	Asia, Pacific
C	adr/ayr	Asia, Pacific
C	adrq-	Asia, Pacific
D	adrq-	Mediterranean
D	ayw2/3	India
E	ayw4	West Africa
F	adw4q-	America
G	Adw2	America, Europe

Table1. Geographical distribution of HBV subtypes versus genotypes.

2.2.1.3 *HBV mutants*

The accumulation of viral mutants depends on their generation rate and on the advantage granted to the virus. HBV, in spite of high replication efficiency with daily production of 10^{11} circulating virus particles per day, shows the rate of mutant generation lower than 2×10^4 base substitutions/site/year (Girones et al, 1989). The structure of the HBV genome with multiple overlapping ORFs reduces the number of viable mutants and the rate of their production. Mutations occur randomly along the HBV genome, and the selection of one over the others warrants a biological advantage to the prevalent mutation during the replication cycle of the virus or a selective advantage to the mutant over wild type virus in host-virus interactions (Oldstone et al, 1991).

2.2.1.3.1 Surface gene mutants.

The envelope proteins of HBV are targets of both humoral and cellular immune responses that are involved in viral clearance, and anti-HBs antibodies show neutralising activity. As a consequence, envelope proteins have been used in the prophylaxis of HBV infection for preparation of the vaccines. However, HBV infection in children after receiving full vaccination has been described.

Most of the described vaccine escape cases had a mutation from glycine to arginine at aminoacid positions 145 of the HBsAg, within the second loop of the α determinant (Carman et al, 1990). This mutation has also been found in liver transplant patients given monoclonal anti- HBs (McMahon et al, 1992). Another mutation in residue 141 associated with breakthrough infection has been described in The Gambia (Fortuin et al, 1994). Both these mutations occur at the second loop of the α determinant (see figure 1). Other mutations have been seen in the α determinant in positions aa 124, 129, 131, 137, 140

(McMahon et al, 1992), but they do not seem to have clinical importance, though substitutions T126S and Q129H have also been shown to exhibit various degrees of altered binding of HBsAg to several monoclonal antibodies (Chiou et al, 1997; Coleman et al, 1999; Seddigh-Tonekaboni et al, 2000). There is a limited information about the prevalence of these escape mutations in the population at large. It is not clear whether the changes arise during the course of breakthrough infections or whether the original inoculum consists of a pool of HBVs that includes the HBV variant. A recent study from Taiwan showed that after mass vaccination the prevalence of α determinant mutants in HBV-DNA positive vaccine escape children increased from 7,8% in 1984 to 28,1% in 1994 (Hsu et al, 1999).

Recently vaccine escape mutants with changes located outside the α determinant have been identified in 6 immunised Singapore infants (Chong-Jin et al, 1999). The significance of these mutants is suggested by the decreased binding of some of them to α determinant-specific monoclonal antibodies.

Detection of HBV unable to produce pre-S2 protein is not a rare event in chronic HBsAg carriers. Double nucleotide mutation in the start codon of the PreS region prevents the synthesis of the protein (Pollicino et al, 1995). One study also showed frequent association of pre-S2 defective virus with fulminant hepatitis (Pollicino et al, 1997).

Truncated integrated pre-S2/S gene sequences have a transcriptional transactivator function and appear to be associated with the development of HCCs (Kekule et al, 1990).

A few patients have been identified who have mutations in the S gene sufficient to prevent normal production of HBsAg despite detectable DNA levels. Insertions of six and nine nucleotides that prevented HBsAg production were also reported (Hou et al, 1995).

In immunosuppressed virus carriers deletions in the pre-S1 region removing the pre-S2 initiation codon and creating a premature termination codon in the pre-S1 and S region were described (Preikschat et al, 1999).

2.2.1.3.2 Mutations in core promoter regain

The core promoter region of the HBV genome regulates transcription of the pre-core and pregenomic mRNAs encoding HBeAg and HBcAg. The nt1762 A to T and nt1764 G to A double mutation in the basal core promoter (BCP) of the HBV genome appears to arise during chronic infection (Kaneko et al, 1995; Kidd-Ljunggren et al, 1995; Laskus et al, 1995) and to become the dominant quasispecies, indicating that there may be a selective advantage for it over wild type viruses, possibly due to enhanced viral replication. This mutation has been observed regardless of the HBeAg status and clinical presentation, in chronic infection and has been associated with fulminant hepatitis (Laskus et al, 1995; Nagasaka et al, 1998) and hepatocellular carcinoma (Baptista et al, 1999).

Further mutations in the core promoter region affecting nt 1766 C to T and nt 1768 T to A were described in a cluster of cases of fulminant hepatitis B. These mutations led to a 15-fold increase in replication as a result of enhanced viral encapsidation of pregenomic RNA into the core particles (Baumert et al, 1998). These mutations were found also in 12 out of 14 Chinese patients with HCC (Fang et al, 1998). One patient had a deletion in this region.

2.2.1.3.3 Pre-core and core mutants and deletions

The presence of HBeAg in the serum of hepatitis B surface antigen- positive patients is conventionally considered to be an indicator of active viral replication and indicate high levels of infectivity. Pre-core mutations that abolish HBeAg expression without a concomitant cessation in viral replication have been described. A guanine to adenine transition at nucleotide 1896 of the pre-core region, resulting in translation stop codon, is the mutation that most often prevents HBeAg expression (Carman et al, 1989). This mutation is present in almost all patients with anti HBe-positive HBV DNA-positive chronic hepatitis in Italy, Greece, Bulgaria, Spain and Japan and is thought to be associated with more severe disease (Dienes et al, 1995). Less often it has been described in North America, France, UK and China. Pre-core mutations have been associated with fulminant hepatitis. Epidemics of fulminant hepatitis caused by hepatitis B virus pre-core mutants have been described (Liang et al, 1991; Sterneck et al, 1998), but other similarities of these viral strains in different epidemics have not been observed, so it is difficult to say that certain mutants cause fulminant hepatitis. *De novo* infection with variant showing a defect in HBeAg and pre-S2 protein synthesis, but not a high replication competence has been described in fulminant disease in infants (Sterneck et al, 1998).

There are no data available on prevalence of pre-core mutations in West Africa, but South African study suggests a large variety of HBV genomes in Africa (Kramvis et al, 1997). Out of 57 HBV-DNA positive samples (42 HBeAg-negative, 15 HBeAg-positive) 10 carriers had virus with mutations preventing HBeAg synthesis. The 1896 stop codon mutation occurred alone only in 1 HBeAg negative carrier. One HBeAg-negative and 3 HBeAg-positive patients carried mixed populations of wild type virus and 1896 mutant. Mutations of the pre-core initiation codon (1814-1816) were found in 4 HBeAg negative

carriers and 1 HBeAg-negative carrier had a nonsense mutation at position 1874. It remains possible that these findings were associated with viral genotype. Genotype A is common in Africa and it is known not to be associated with the 1896 stop codon mutations. In genotype A, the 1896 mutation coexists with a second mutation (C->T at position 1858 in codon 15), and both mutations are paired in the secondary structure of the RNA encapsidation signal, which can explain the rare presence of pre-core mutants in this genotype (Rodriguez-Frias et al, 1995).

HBe mutations are found rarely in HBeAg positive patients with asymptomatic disease. Changes in amino acid sequences away from consensus are most frequent in anti-HBe positive persons or patients with severe chronic or fulminant hepatitis. Core genes in HBsAg-positive Japanese chronic hepatitis patients had more amino acid substitutions or deletions than those without disease (Ehata et al, 1992).

The core gene possesses both humoral and cytotoxic T-cell (CTL) epitopes (Bertoletti et al, 1991). Mutations within a cytotoxic T-cell epitope of the viral nucleocapsid protein have been shown to abrogate or antagonise the T-cell response to HBV and lead to viral persistence (Bertoletti et al, 1994). In cross-sectional studies, representing accumulated variability over time and many different clinical situations, mutations were found significantly more frequently in known HBe T helper cell and B cell epitopes (Carman et al, 1995) than in interepitope regions, implying that they can be result of immune selection or have some growth advantage. New dominant variants with changes within B and CD4+ cell epitopes occurring during the course of disease were detected during occasional exacerbations in a chronic carrier (Alexopoulou et al, 1997). In another follow up study of anti-HBe seroconverters who went into remission, mutations occurred in the CD4+

epitope, but for patients with ongoing disease mutations were most likely to be found in B cell epitopes (Carman et al, 1997). In those who remit, mutations in the major CD4+ cell epitope may allow immune escape, thus minimising immune-mediated hepatitis (Carman et al, 1997). Whether mutations in B and Th cell epitopes are a cause or an effect of disease are entirely unresolved.

Deletions in the HBV core gene have been described and often were associated with severe liver damage (Ehata et al, 1992). HBV variants with core gene deletions are thought to inhibit HBV replication, do not persist in preference to wild-type HBV and do not confer resistance to interferon-alpha treatment (Marinos et al, 1996). Recently, a number of deletions in core region were described in immunosuppressed patients (Preikschat et al, 1999).

2.2.1.3.4 Polymerase gene mutants

The HBV genome encompasses four overlapping ORFs with the polymerase gene overlapping the envelope gene in its entirety. As a result, mutations in the catalytic domain of the polymerase gene can affect the aminoacid sequence of the HBsAg and vice versa. In particular, the genetic sequence for the *a* determinant (Carman et al, 1997), actually overlaps the major catalytic regions of the viral polymerase protein from amino acid 454 to 524 and known as domains A and B (Aye et al, 1997). Surface gene mutants associated with mutations in the polymerase gene after treatment with lamivudine or famciclovir have been described.

2.2.2 Molecular evolution of HBV genome

2.2.2.1 Evolution in different populations and phylogenetic analysis

The evolutionary origins of HBV are unclear. The long infectious period and transmission routes do not require large host populations to establish an infection. It has been suggested that the divergence of the viral genotypes may reflect the migration of human population over the last 100000 years (Norder et al, 1994).

The most recent studies on the origin of HBV have employed the phylogenetic tree approach. Analysis using the neighbour-joining method and the unweighted pair-grouping method was done on 27 samples from different HBV subtypes (Mizokami et al, 1997). Phylogenetic trees for five regions: PreS, S, X, C, and P were constructed. The most frequent synonymous substitutions per nucleotide site were observed in the non-overlapping parts of the P and C genes. The hypothesis that the evolution of HBV was constrained by the overlapping genes was suggested, and thus the rate of evolution of HBV is slower than in other viruses.

Similar observations were made earlier by another group (Yang et al, 1995), using the maximum likelihood method on 13 complete HBV genomes. The variation in nucleotide conservation was not random around the HBV genome. Interestingly, genetic relatedness between serotypes was not observed.

A New World origin of HBV was first suggested in 1997 (Bollyky et al, 1997). A subsequent study using phylogenetic analysis with maximum-likelihood (ML) and neighbour-joining (NJ) bootstrap analysis on 39 complete genome sequences supported this hypothesis (Bollyky et al, 1999). It also showed, that African primates may have acquired infection from humans.

Phylogenetic trees can also be used to trace the dynamics of viral transmission within small populations. Several point source outbreaks of other viral infections have been investigated using these methods. For example, phylogenetic tree analysis of HCV sequences obtained after transmission of anti-rhesus D immunoglobulin (Ig) in Ireland provided conclusive evidence that certain infected batches of the preparation were responsible for the cluster of cases seen (Power et al, 1995). An epidemiological investigation indicated that six patients treated in a haematology unit who developed acute hepatitis B may have been infected as a result of contamination of a liquid nitrogen bone marrow storage tank. This hypothesis was supported by analysis of nucleotide sequences (Hawkins et al, 1996). In another study, phylogenetic trees of HIV sequences were used to show that a group of patients was possibly infected by a dentist in Florida (Ou et al, 1992). As far as I know, phylogenetic studies on outbreaks of HBV infection have not been done.

It is well known that the Pre-core/Core regions are the most variable in the HBV genome. In this respect, phylogenetic analysis of these regions would be most appropriate. ML and NJ bootstrap analysis are the currently recommended tools for phylogenetic analysis (Bollyky et al, 1999).

2.2.2.2 Evolution of the genome within the host

Unfortunately, very few longitudinal studies of molecular sequences in HBV carriers have been reported so far. The limited data that is currently available suggest that most of the mutations occur during periods of increased immune pressure, around the time of HBeAg seroconversion and during exacerbations of chronic hepatitis.

During chronic HBV infection, patients with HBeAg frequently seroconvert to anti-HBe (Realdi et al, 1980). This has been associated with the occurrence and prevalence of precore-defective mutants. Several years are required for complete replacement of wild-type viruses by mutants (Aye et al, 1994; Uchida et al, 1994). A study of 26 Japanese HBV carriers showed that the number of amino acid changes in the core region in HBeAg negative carriers (5.9 ± 3.8) was significantly higher than that in the HBeAg positive carriers (1.5 ± 1.0) (Karasawa et al, 1997). 95% of HBeAg negative samples in this study contained pre-core stop codon mutations. Another group in Chinese patients reported similar findings over a period of four years (Bozkaya et al, 1996).

In HBeAg positive samples a very high degree of conservation of the pre-core/core region can be maintained for a long time period. Completely identical nucleotide sequences were present in Chinese family members over a 20-year period and after perinatal transmission (Bozkaya et al, 1997). Similar observations were made on Vietnamese and Turkish families (Hannoun et al, 2000). Some sequences did not change over 20-30 years. In one out of seven Japanese patients followed up for 7 years, the virus sequence remained the same during this time (Uchida et al, 1994).

2.2.3 Vaccines

Prevention of HBV infection is the most effective weapon for its control.

The first licensed hepatitis B vaccine consisted predominantly of purified 22-nm particles derived from plasma of chronic carriers. This vaccine was shown in a controlled trial in American homosexual men to be at least 95% effective (Szmuness et al, 1980).

The limited availability of infected human plasma and the possible risk of HIV infection led to the development of a recombinant vaccine prepared from antigen expressed in the yeast *Saccharomyces cerevisiae*. This vaccine was shown to be equally effective (Stevens et al, 1987).

Currently both types of vaccine are available and plasma vaccine cost is down to 60 cents per injection. Vaccination against hepatitis B is shown to be cost effective in prevention of HCC and chronic liver disease and it is recommended by WHO.

New DNA vaccination approaches are still under development for both prevention and treatment of chronic HBV infection (Encke et al, 1999).

2.3 Transmission of hepatitis B virus

2.3.1 Modes of transmission

Percutaneous exposure to blood, sexual transmission, perinatal transmission and horizontal transmission through close physical contact account for almost all HBV infections in humans. Studies in animals support these observations. Primates have been experimentally infected with HBV by inoculation of blood, semen, and saliva (Alter et al, 1977; Bancroft et al, 1977}. Intranasal or peroral administration of saliva failed to infect chimpanzees and

gibbons. Infection by faeces, urine, tears and breast milk has never been demonstrated although HBsAg or HBV particles have been detected in such fluids

2.3.1.1 Percutaneous and parenteral transmission

Transmission of the infection may result from accidental inoculation of small amounts of blood or contaminated fluid during medical, surgical or dental procedures, immunisation with inadequately sterilised syringes and needles (Seeff et al, 1987), intravenous and percutaneous drug abuse, tattooing, ear piercing, and accidental inoculation with razors that have been contaminated with blood. Infected health care workers can also be a significant source of infection (Rimland et al, 1977).

Before blood screening was introduced, in developed countries disease was thought to be acquired mostly by transfusion of blood or other contaminated blood products. Prevalence rates of hepatitis B were highest in haemophilic patients who previously received blood-clotting components.

2.3.1.2 Sexual transmission

HBV infection in spouses of carriers has been described (Henigst et al, 1973; Berris et al, 1973). In one study six (18%) of 34 spouses of HBsAg positive carriers observed for at least 4 months developed clinical hepatitis, and asymptomatic infection was detected in an additional 16% followed for 7 months or longer. In only two of 136 other family members subclinical infection was detected (Mosley et al, 1975). Exposure to infected semen and menstrual blood during sexual intercourse is suspected to be the source of infection (Alter et al, 1977; Inaba et al, 1979).

The sexually promiscuous, particularly male homosexuals are thought to be at higher risk of infection with HBV (Fulford et al, 1973; Coleman et al, 1977; Szmuness et al, 1975). In a study of 3816 homosexual males who attended sexually transmitted diseases clinics, 61.5% were positive for serological markers and 6% became chronic carriers, a significantly higher rate than in other group of patients (Schreeder et al, 1982).

There are remarkably few studies of sexual transmission in Africa. However, since most of the population is infected in early childhood, it may play a limited role in overall transmission rates. A study in Tanzania has shown that sexual transmission occurs in 7.2% in men and 3% in women if the partner was known to be a carrier (Jacobs et al, 1997).

2.3.1.3 Maternal- neonatal transmission

Perinatal exposure to maternal blood is highly efficient mode of HBV transmission. During the perinatal period infants are exposed to maternal blood through placental tears, trauma related to birth and contact of conjunctiva and mucous membranes with blood and other fluids during labour and delivery.

The incidence of perinatal transmission depends primarily on the prevalence of carrier mothers and the risk of transmission of the virus from a carrier mother to her infant (Edmunds et al, 1996). The risk of perinatal transmission is far greater from HBeAg positive mothers than from HBeAg negative mothers (Beasley et al, 1977; Schweitzer et al, 1973; Tong et al, 1981).

In Eastern Asia around 70-85 % of children born to HBeAg positive mother become HBsAg positive during the first 3 months after birth (Beasley et al, 1977). A greater proportion of carrier mothers in east Asia appear to be HBeAg positive and have higher levels of circulating HBV DNA compared to African mothers (Greenfield et al, 1986). Even when mothers in Africa are HBeAg positive, their babies do not become HBsAg positive until 6 months to 1 year later (Marinier et al, 1985).

2.3.1.4 Horizontal transmission

Most hepatitis B virus infections in sub-Saharan Africa are acquired through horizontal transmission in early childhood (Whittle et al, 1983; Tabor et al, 1985), but the exact mechanisms of spread have not been documented. Study in North America found an association between intra-familial transmission and skin sores (Bernier et al, 1982). In West Africa virus was found in exudates from tropical ulcers (Foster et al, 1984).

The possibility of transmission by blood-feeding arthropods has been of concern but never proven. Experimental feeding of HBV- contaminated blood to mosquitoes revealed that the virus was digested within 48 hours. Because mosquitoes tend to feed every 72 hours transmission this is unlikely. Bedbugs retain virus up to 30 days after a single blood meal (Ogston et al, 1979) and virus is detectable in their faeces for up to 6 weeks (Ogston et al, 1980).

The hypothesis of bedbug transmission was tested in The Gambia. Risk factors for hepatitis B virus transmission were examined in 973 Gambian children aged 6 months to 5 years. Skin disease and presence of bedbugs in the bed were found to be associated with

transmission (Vall Mayans et al, 1990). Later, a controlled trial of bedbug elimination in several villages failed to reduce the risk of HBV infection (Vall Mayans et al, 1994).

In a large study enrolling 1385 persons performed in Ghana (Martinson et al, 1998), the behaviour most strongly associated with prevalence of HBV was sharing of bath towels, sharing of chewing gum or partially eaten candies, sharing of dental cleaning materials and biting of fingernails in conjunction with scratching the backs of carriers.

Horizontal transmission is known to be common in tropical parts of South America, where the environment is quite similar to that in Africa. In a study in Brazil of families that were moved 1500 km to the Amazon region, a significant rise in acute infections was described soon after arrival. An incidence rate of 7.2 new infections per 100 exposed subjects per month was found (Dutra et al, 1998). All the main known transmission factors such as tattooing, ear piercing, circumcision, surgical or dental procedures, immunisation with inadequately sterilised syringes and needles were excluded and unknown environmentally related factors were suggested.

Recently, in the study on 54 chronic HBsAg carriers HBV DNA was found in urine in 82% of HBeAg positive patients and in 24% of HBeAg negative patients (Knutsson et al, 2000). Therefore urine should be regarded a potential route of transmission and therefore be investigated further as a means of horizontal transmission of HBV.

2.4 Natural History of HBV infection

2.4.1 Clinical presentations

2.4.1.1 Acute hepatitis

Hepatitis B infection can be divided into acute and chronic hepatitis. The course of *acute hepatitis B* can be divided into four clinical phases: (a) the incubation period; (b) a prodromal or preicteric stage; (c) the icteric phase and the convalescent period (Fields BN, 1997).

Usually the incubation period ranges from 45 to 120 days (Mosley et al, 1975). The incubation period can be influenced significantly by the size of inoculum (Allen et al, 1970). Coinfection with other hepatitis viruses, chemical or physical factors, age, immunosuppression or other unusual viral- host interactions can shorten the incubation period. Late administration of specific antibody in some cases is not protective but can delay the development of clinical disease (Grady et al, 1975).

In symptomatic patients a short prodromal and preicteric phase, varying from several days to more than a week, precedes the onset of jaundice in more than half of the adult patients. Usually it is characterised by mild fever, easy fatigability, malaise, anorexia, myalgia, nausea and vomiting. Hepatomegaly usually precedes jaundice and patients can feel pain or discomfort in the right upper quadrant. The icteric phase usually starts with appearance of dark urine due to bilirubinuria followed by pale stools and jaundice, which becomes clinically apparent when the total bilirubin level exceeds 2 to 4 mg/dl. On physical examination the liver is usually enlarged. Usually clinical symptoms subside in a few months and full clinical recovery occurs. Clinical signs associated with poor prognosis, are

rapid decrease of a liver size, prolongation of protrombin time to greater than 50 seconds, an elevated bilirubin level of greater than 17,5 mg/dl, often a presence of a rapidly declining ALT value and encephalopathy developing within seven days of the onset of jaundice (O'Grady et al, 1989). The severity of disease is proportional to increasing age of the patient. Infants usually have asymptomatic disease.

Fulminant hepatitis has been designated as acute hepatitis which in first 8 weeks of illness leads to severe impairment of hepatic synthetic processes and other functions causing sudden onset of high fever, abdominal pain, vomiting, and jaundice, followed by hepatic encephalopathy associated with coma and seizures (Berk et al, 1978). Ascites, bleeding, renal dysfunction and decerebrate rigidity lead to death in more than half of the patients. Mortality is highly correlated with increasing age.

2.4.1.2 Chronic hepatitis

Some chronically infected patients may have no clinical or biochemical evidence of liver disease. This group is often categorised as asymptomatic hepatitis B carriers or simply HBsAg carriers. Disease can be histologically classified as chronic persistent, chronic active, and chronic active hepatitis with cirrhosis, but this classification is rather approximate and often does not reflect the dynamics of the disease.

The main method of classifying chronic viral hepatitis is by cause and the old histology-based classification is no longer considered appropriate. The concepts of grading and staging, borrowed from tumour pathology, have been introduced (Scheuer, 1996).

Patients with moderate to severe chronic hepatitis may have no symptoms, but may also be significantly incapacitated. Easy fatigability, anxiety, poor sleep, anorexia, and malaise are

the most common complaints. During follow-up, a series of remissions and relapses are usually observed. Remissions may last a few months or years. Patients may have jaundice and elevated transaminases, especially during the relapse of disease. Ascites and pedal oedema may be present. Encephalopathy or variceal bleeding can be occasional presentations of the disease and usually this is associated with progression to cirrhosis.

Survival rates at 5 years in chronic persistent hepatitis are thought to be around 97% of patients with chronic persistent hepatitis, 86% for those with chronic active hepatitis and 55% for those with active hepatitis and cirrhosis (Weissberg et al, 1984). Age is the most important factor for survival. Prognosis for the elderly is worse because of decreased function of other organ systems.

2.4.2 Pathogenesis and pathology

Both CD8⁺ and CD4⁺ cells are thought to be crucial in immune response against HBV. Patients who successfully clear acute HBV infection characteristically mount a strong, polyclonal, multispecific CD4⁺ and CD8⁺T lymphocyte responses to several epitopes in the HBV envelope, nucleocapsid and polymerase proteins (Ferrari et al, 1990; Ferrari et al, 1991; Jung et al, 1994, Rehmann, 1995). HLA class I restricted CTLs have a key role in the control of HBV infection by destroying infected cells. This process leads to direct cell killing by the CD8⁺ cytotoxic T lymphocytes (Chisari et al, 1995). The polymorphic nature of the MHC binding sites and differences in the T-cell repertoire among people leads to highly variable binding affinity for the immunodominant HBV peptides, which can determine outcome after acute HBV infection (Carman et al, 1997). The different immune responses in patients in whom the virus is cleared successfully and those, in whom it is not, depend on the match between the HBV peptides presented by host MHC

molecules and the specific T-cell repertoire of the host. Highest frequency of circulating HBV-specific CD8 cells was also shown to coincide with the clinically acute phase of hepatitis (Maini et al, 1999). If there is sufficient recognition and activation, the immune response is carried to completion, infected cells are destroyed, viral replication is aborted and antibodies to HBsAg prevent the reinfection of hepatocytes are produced. Some studies indicate that long lasting immune response can be maintained by traces of virus that persist indefinitely after clinical and serological recovery (Penna et al, 1996). Following this reasoning, HBV-specific CD8⁺T lymphocytes contribute to chronic inflammatory liver disease by destroying some of the infected hepatocytes, but they remain quantitatively ineffective in eliminating infected hepatocytes or mediating viral clearance. CD8 cell response differs in patients with active liver disease and active viral replication and in patients without liver inflammation and HBV replication. High frequency of intrahepatic HBV specific CD8⁺ cells was found in the liver of patients from both groups, but in patients with active disease these cells were more diluted among liver infiltrates of other non-specific cells (Maini et al, 2000). Inhibition of HBV replication was also associated with the presence of a circulating reservoir of CD8⁺ cells able to expand after specific virus recognition that was not detectable in highly viremic patients with liver inflammation.

Some other important mechanisms of clearance of infection have been suggested. T lymphocytes may directly inhibit viral gene expression and replication and thus inactivate HBV without killing the infected hepatocytes (Cavanaugh et al, 1997; Chisari et al, 1997; Guidotti et al, 1997). HBV DNA was even shown to nearly disappear from the liver and the blood of acutely infected chimpanzees long before the peak of T cell infiltration (Guidotti et al, 1999). That has lead to suggestion that HBV infection can be primarily

controlled by a noncytopathic, cytokine dependant curative mechanism, named viral purging (Guidotti et al, 1999). This effect was found to be mediated by gamma interferon (INF- γ), tumor necrosis factor alpha (TNF- α) and other interferons in transgenic mice model (McClary et al, 2000). The cytokine-inducible downstream events that exactly inhibit HBV replication have not been well defined. Nitric oxide (NO) was shown to mediate the ability of INF- γ but not IFN- α/β to inhibit HBV replication (Guidotti et al, 2000).

The pathology in HBV infection usually does not differ from the pathology of hepatitis from other causes. It is difficult to distinguish acute hepatitis from exacerbation of chronic active hepatitis. Early events in viral hepatitis are not well described, but occasionally an increase in liver-cell mitoses or a proliferation of sinusoid-lining cells is seen.

Fully developed hepatitis is characterised by degeneration and death of liver cells accompanied by proliferation of Kupffer cells and an infiltrate composed mainly of mononuclear cells. Portal tracts are the main sites of such infiltrates. The liver-cell changes are apparent throughout the lobule, but vary in intensity and are most severe in centrilobular areas. Liver cells are commonly swollen (ballooning degeneration), but they can shrink (acidophilic change) (Bianchi et al, 1970).

In the later stage of evolution of the lesion there is often a variable degree of collapse and condensation of reticulin fibres, and accumulation of ceroid pigment and stainable iron in large phagocytic cells. Portal-tract enlargement may lead to linking of portal tracts without interference with lobar architecture.

In chronic disease the portal and periportal alterations are often more prominent. When confluent necrosis links vascular structures, it is called *bridging necrosis*. *Piecemeal necrosis* is the death of hepatocytes at an interface between parenchyma and the inflamed connective tissue of a portal tract or fibrous septum.

Immunohistochemical staining is often helpful. The presence of HBsAg in hepatocytes is an indicator of chronicity in HBV infection, and HBcAg in these cells is a marker of active viral replication, especially when it is abundant and located in hepatocyte cytoplasm as well as in nuclei (Scheuer et al, 1996).

The stage of chronic hepatitis is related to its time course. Its histological evaluation is based on the extent of fibrosis and development of fibrosis. Connective tissue stains are essential for staging. In chronic hepatitis, fibrous tissue is deposited in and around portal tracts, usually in association with periportal necroinflammatory activity. Perihepatocellular fibrosis may lead to the formation of hepatocyte rosettes. Extensive piecemeal necrosis may extend to portal tracts and result in development of portal-portal septa. Then fibrous septa extend from the portal tracts into hepatic lobules for varying distances and eventually reach the central hepatic venules. Portal septa are thought to be more important in the development of cirrhosis.

Cirrhosis is the final and irreversible stage of chronic hepatitis. It is characterised by presence of parenchymal nodules, which are surrounded by fibrous septa. This leads to functional alterations of blood flow with increase of portal pressure.

2.4.3 Serological findings during HBV infection

A number of serological assays are available to differentiate acute from chronic or past infection. In an acute case HBsAg is present for some days to 2 weeks before the onset of symptoms and is at maximum titre at the height of liver damage. In most cases HBsAg cannot be detected beyond 3-4 months. In 5-10% of patients antigenemia is of short duration and may not be detectable at the onset of clinical symptoms. In such cases the presence of antibodies, especially IgM anti-HBc, and anti-HBe may be the only indications of HBV infection.

Methods such as reverse passive haemagglutination (RPA) and enzyme immunoassay/radioimmunoassay (EIA/RIA), are highly sensitive and specific and allow detection of HBsAg at 100-200 pg/ml serum, that is, about 3×10^7 particles/ml (Schmilovitz-Weiss et al, 1993). These assays are used most commonly for the detection HBsAg in serum and plasma. In RPA, fixed erythrocytes coated with anti-HBs are added to test samples, and haemagglutination patterns are read. In RIA a sandwich method, with anti-HBs as both absorbed reagent and as tracer have been employed. RPA is marginally less sensitive than the other two serological methods on Gambian elderly but not in young children (Whittle, personal communication).

The immunoglobulin class of anti-HBc can differentiate between acute and chronic HBV infection. Patients with acute hepatitis develop high titres of IgM-specific anti-HBc antibody commencing with aminotransferase abnormalities (Cohen et al, 1978; Lemon et al, 1981). These titres gradually decline over a 6- to 8- month period regardless of whether the disease resolves or becomes chronic. Low concentrations of IgM-specific anti-HBc have been detected up to 2 years in patients whose disease has resolved successfully and in

many chronic carriers (Sjogren, 1983; Tsuchie, 1984; Tsuda, 1984). Among HBsAg carriers the predominant component is a 7S IgM fraction versus a 19S component in acute disease. Fewer than 15% of chronic hepatitis B patients have an IgM-specific anti-HBc titer that is high enough to be detected in the conventional IgM anti-HBc assay; but the reactivity of the test is relatively low when compared with that observed in acute disease. IgM anti-HBc is detectable during reactivation of chronic disease or especially after steroid withdrawal. So the presence of IgM-specific anti-HBc at a dilution of 1:1000 or higher is presumptive evidence a recent acute infection.

Detection of anti-HBc and anti-HBc in HBsAg negative specimens normally indicates a previous infection and confirms immunity against reinfection. However such serological results can be observed in acutely ill neonates and in those with fulminant hepatitis in which HBsAg expression by necrotic liver tissue is severely restricted. Then, IgM specific anti-HBc can be very helpful.

Anti-HBs is the principal neutralising antibody of hepatitis B, and its appearance usually signifies recovery from infection. Patients with pre-existing anti-HBs are usually protected against HBV infection. Detection of anti-HBs alone follows exposure to HBsAg after vaccination.

To determine whether the blood or body fluids of an HBsAg positive patient are highly infectious or contain a relatively small concentration of infectious virions, assays for

detection of HBeAg, anti-HBe, or HBV DNA can be used. Most acutely infected patients are HBeAg positive. HBeAg positive specimens contain high concentrations of infectious particles and HBV DNA (van Ditzhuijsen, 1985). The presence of HBeAg shows active

viral replication. In contrast, in anti-HBe positive samples, the number of hepatitis B virions is markedly reduced. However, a subset of anti-HBe- positive carriers may be HBV DNA positive, especially those with pre-core mutations. The clinical course of the disease can be very aggressive. HBeAg testing for HBsAg positive pregnant women identifies those of them, whose infants are at risk of becoming infected.

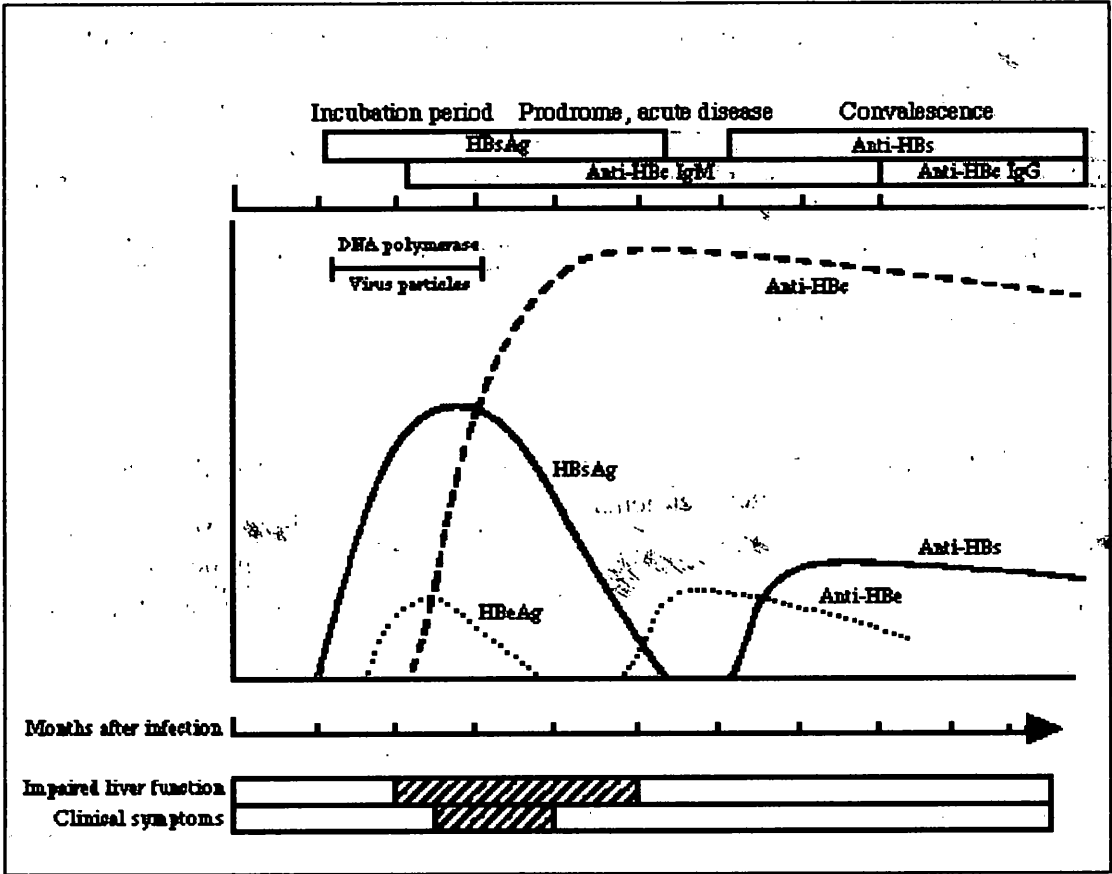


Figure 2. Serological findings during the typical course of acute HBV infection.

Stage of the infection	HBsAg	HBeAg	Anti-HBe	Anti-IgM HBc	Anti-IgG HBc	Anti-HBs
Incubation period	+	+	-	-	-	-
Acute hepatitis	+	+	-	+	+	-
Chronic carriage	+	+/-	+/-	-	+	-
Recovery	-	-	-	-	+	+/-
Immunisation	-	-	-	-	-	+

Table 2. Serological findings for the different stages of HBV infection.

2.4.4 Polymerase chain reaction and sequencing.

Detection of HBV DNA is a useful tool for evaluating virus load in serum and assessing efficacy of antiviral treatment. The most sensitive assay for HBV DNA detection is the polymerase chain reaction (PCR) (Carotenuto et al, 1995).

PCR products can be cloned and sequenced afterwards or sequencing can be done directly. Cloning and sequencing could give much better insight into the presence of quasispecies in the sample. It is widely accepted that sequencing of PCR product directly will reveal the dominant sequence (Akarca et al, 1994). From longitudinal studies, it has become apparent that by direct sequencing of PCR products it is difficult to tell whether changes in sequences represent true mutations or changes in the proportion of different quasispecies.

2.4.5 Levels of aminotransferases as a diagnostic test

Alanine-aminotransferase (ALT), an enzyme present in liver cells, with lesser amounts in the kidneys, heart, and skeletal muscles, is a relatively specific indicator of acute liver cell damage. When such damage occurs, ALT is released from the liver cells into the bloodstream, often before jaundice appears, resulting in abnormally high serum levels that may not return to normal for days or weeks.

ALT levels by a commonly used method range from 10 to 40 U/L. The normal range for infants is twice that of adults. Very high ALT levels (up to 50 times normal) suggest viral or severe drug-induced hepatitis, or other hepatic disease with extensive necrosis (death of liver cells). (AST levels are also elevated but usually to a lesser degree.) Moderate-to-high levels may indicate infectious mononucleosis, chronic hepatitis, intrahepatic cholestasis or cholecystitis, early or improving acute viral hepatitis, or severe hepatic congestion due to heart failure. Slight-to-moderate elevations of ALT (usually with higher increases in AST levels) may appear in any condition that produces acute hepatocellular (liver cell) injury, such as active cirrhosis, and drug-induced or alcoholic hepatitis. Marginal elevations occasionally occur in acute myocardial infarction (heart attack), reflecting secondary hepatic congestion or the release of small amounts of ALT from heart tissue.

Aspartate-aminotransferase's (AST) levels measured by a commonly used method range from 8 to 20 U/L although some ranges may express a maximum high in the 40s. AST levels fluctuate in response to the extent of cellular necrosis (cell death) and therefore

may be temporarily and minimally elevated early in the disease process, and extremely elevated during the most acute phase. Depending on when the initial sample was drawn, AST levels can rise- indicating increasing disease severity and tissue damage- or fall- indicating disease resolution and tissue repair. Thus, the relative change in AST values serves as a reliable monitoring mechanism. Maximum elevations are associated with certain diseases and conditions. For example, very high elevations (more than 20 times normal) may indicate acute viral hepatitis, severe skeletal muscle trauma, extensive surgery, drug-induced hepatic injury, and severe liver congestion. High levels (ranging from 10 to 20 times normal) may indicate severe myocardial infarction (heart attack), severe infectious mononucleosis, and alcoholic cirrhosis. High levels may also occur during the resolving stages of conditions that cause maximal elevations. Moderate-to-high levels (ranging from 5 to 10 times normal) may indicate chronic hepatitis and other conditions. Low-to-moderate levels (ranging from 2 to 5 times normal) may indicate metastatic hepatic tumours, acute pancreatitis, pulmonary emboli, alcohol withdrawal syndrome, and fatty liver (steatosis).

2.5 HBV and HCC

The geographical distributions of HCC and the HBV carrier state, at both global and local levels, are remarkably similar (Szmuness et al, 1978). Even though carriers may represent 5% to 10% of the population, they account for 50% to 80% of the HCC patients in endemic areas. The size of the risk is best illustrated by the cohort study on 3454 Taiwanese carriers of the virus which showed that 5% of these carriers died from HCC (Beasley et al, 1981). HBV carriage was associated with a 100-fold increase in the risk for HCC development in relation to non-carriers (Beasley et al, 1988). Patients with HCC had

a significantly higher prevalence of markers of current HBV infection in serum and liver tissue than matched controls (Szmuness et al, 1978). HBV carriage usually precedes the onset of HCC by many years; an interval consistent with a cause and effect relationship, and the chance of neoplastic transformation rises progressively with increasing duration of the infection. From the age distribution of Taiwanese hepatoma cases, it can be inferred that most tumours arise after 30 or more years of persistent infection; relatively few cases occur in children or adolescents. Sequences of viral DNA are integrated in most of the HBV related HCC cancers.

Woodchuck hepatitis virus induces HCC in eastern woodchucks with a frequency of 100% within 2 years of infection, whilst 30% of ground squirrels infected with ground squirrel hepatitis virus develop HCC within 6 years (Popper et al, 1987).

Not only HBsAg carriage but also presence of antibodies to HBsAg after infection was found to be associated with a 4,7-fold increased risk for HCC (Yu et al, 1997) in non-Asians of Los Angeles, but other researchers have never repeated these observations.

2.6 Chronic HBsAg carriage

2.6.1 Definition

A chronic HBV infection is defined as the presence of HBsAg in serum for 6 months or more. This definition is based on a study of patients in hemodialysis units whose seroconversion from HBsAg negative to HBsAg positive was observed. In the patients with persistent HBsAg detection for 6 months was associated with a 90% probability of remaining positive indefinitely (London et al, 1977).

2.6.2 Risk factors

2.6.2.1 Influence of age at exposure

The major determinant of chronicity of HBV infection is age at exposure (Edmunds et al, 1993; Coursaget et al, 1987; McMahon et al, 1985). In vertically acquired infection, chronic infection establishes in 75-90% of cases (Beasley et al, 1983). In horizontal transmission in early childhood, chronic infection occurs in 10-15% of infected cases (Ryder et al, 1992). When infection is acquired in adulthood, the rate of chronic infection is around 5% (Nielsen et al, 1971; Tassopoulos et al, 1987).

2.6.2.2 Concurrent infection.

A large increase was observed in the rate of persistent infection in homosexual men who acquire HBV infection in the early phases of HIV infection (Bodsworth et al, 1991; Hadler et al, 1991). Coinfection also may contribute to higher levels of HBeAg expression and a lower degree of inflammation (Goldin et al, 1990).

2.6.2.3 Alcohol

Alcohol is a well-known hepatotoxic agent and may exacerbate HBV-related liver damage through the development of cirrhosis. Persons chronically infected with HBV and who drink alcohol regularly might be expected to be at higher risk of both cirrhosis and HCC (Oshima et al, 1984).

2.6.2.4 Aflatoxin

Exposure to aflatoxin-B1 (AFB1) in several areas of the world has been positively correlated with high mortality rates for HCC. In a prospective study in Shanghai, urinary excretion of aflatoxin metabolites increased the risk of development of HCC fourfold. HBV infection alone increased the risk about seven times, but persons who both excreted aflatoxin metabolites and were HBV carriers had 60 times higher risk than HBV uninfected, AFB1 unexposed individuals (Ross et al, 1992).

2.6.2.5 Human genetic factors

Family associations. The outcome of viral hepatitis infection depends on a number of variables in both the host and the infecting virus. Early studies suggested an autosomal recessive mode of inheritance for chronic HBsAg carriage (Hann et al, 1982), but this seems to be an oversimplification. The same group showed that fathers of patients with chronic HBV infection have significantly lower rates of anti-HBs antibodies after infection. It was suggested that genetic control of anti-HBs production can play a role in development of chronic infection. One report clearly demonstrates that the same virus isolate may lead to two different clinical outcomes in different hosts (Karayiannis et al, 1995). A small study of Chinese twins shows that the outcome of HBV infection is very much influenced by the host genotype (Lin et al, 1989). Monozygotic twins in this study

were chronic carriers in 50% of family cases but dizygotic twins were carriers in 20% of families. No difference was observed in infection rates between both groups.

MHC Class I Associations. Intracellular HBV is eliminated by the CD8+ cytotoxic cells, which are able to recognise viral antigenic peptides presented by MHC class I glycoproteins. In patients with chronic HBV infection the CTL response is decreased (Chisari et al, 1995). Results of MHC association studies comparing the distribution of MHC class I alleles in patients with chronic hepatitis, asymptomatic HBsAg carriers and controls are not very consistent. Moderate and severe hepatitis has been associated with an increased frequency of HLA-B35 and a decreased frequency of HLA-DQw1 (Mota et al, 1987; van Hattum et al, 1987). Asymptomatic carriage has been associated with HLA-B15 (Giani et al, 1979).

MHC Class II associations. Variations in MHC class II amino acid sequences determines which viral antigenic peptide fragments are presented to CD4+ helper cells. Observations show that Th cell responses in patients with chronic HBV infection are reduced in comparison to those in patients with acute infection (Ferrari, 1990). The DR2 locus has been associated with clearance of infection and DR7 with persistent infection in a study performed in Qatar (Almarri et al, 1994). Analysis of MHC class II allele frequencies in The Gambia showed an association of the allele DRB1*1302 with clearance of HBV infection (Thursz et al, 1995). The association has been confirmed in a European study (Hohler et al, 1997).

In patients, who seroconvert from HBeAg- positive to antibody to HBeAg (anti- HBe) after therapy with interferon, Th and CTL responses have increased substantially (Jung et

al, 1995). In a small study DR6 and DR3 alleles were associated with viral clearance after interferon therapy (Scully et al, 1990).

Mannose-binding protein associations. It was discovered that patients with chronic HBV infection have a defect in opsonization that is not seen in other patients with chronic liver disease (Munoz et al, 1987). HBV has a mannose-terminated carbohydrate chain on the middle envelope protein (pre-S2 region), which can be a potential target for mannose-binding protein (MBP). MBP is a known opsonizing protein. Binding of HBV to MBP can enhance the clearance of HBV from peripheral blood (Gerlich et al, 1993). MBP codon 52 'mutant' allele has been found more frequently in patients with persistent HBV infection (Thomas et al, 1996), but this was not confirmed by further more extensive studies {M.Thursz, personal communication}.

TNF- α associations TNF- α being a proinflammatory cytokine may help to control the replication of HBV. Suppression of HBV antigen expression by TNF- α has been demonstrated in HBV transgenic mice (Gilles et al, 1992). TNF- α levels are higher in chronic carriers, who are successfully treated with interferon (Daniels et al, 1990). The rarer TNF2 allele has been shown in functional studies to be associated with increased TNF- α production (Wilson et al, 1993). The same allele is associated with cerebral malaria and may be weakly associated with clearance of HBV infection (Thursz et al, 1996).

Gender differences. It has been shown that males after getting infected in childhood and adulthood are more likely to develop a persistent infection than females and in patients treated for persistent HBV infection females are better responders (London et al, 1977; Realdi et al, 1990}. In an Alaskan population-based study of 1400 carriers, women were significantly more likely to clear HBsAg than men (McMahon et al, 1990). One study comparing sex to clearance of HBeAg showed that clearance was higher in women than in men (McMahon et al, 1985).

In chronic hepatitis the major sequelae, such as fibrosis and cirrhosis, are more common in men than in women. A large prospective Taiwanese study showed that the age-specific incidence of hepatocellular carcinoma was threefold greater in men than in women over the age of 45 (Beasley et al, 1982).

Explanations for the sex difference have not been found, but it is known that replication of HBV is sensitive to sex hormones (Almog et al, 1992). HBsAg clearance by phagocytic cells in inbred mice was achieved more quickly in females than in males (Craxi et al, 1982).

2.6.3 Theoretical background for sample collection for genetic studies in The Gambia.

Recently molecular genetic tools have increased our ability to discover variations (polymorphisms) and correlate them with human disease. Investigators can scan a large number of subjects and identify multiple associated alleles. Approach used on Gambian samples was genome wide scan, where densely distributed markers- microsatellites were used to identify areas of interest in the genome. Distortions in the frequencies of these

markers were compared between a cohort of patients with a specified outcome (e.g. chronic HBsAg carriage) and those without the outcome or in a group of families with the disease phenotype (Dib et al, 1996).

This approach was used successfully previously to detect genetic associations for development of infectious disease- tuberculosis (Bellamy et al, 1998). If a disease-associated polymorphism and a marker are linked, then the marker will be found significantly more frequently in those with the disease. In diseases without previously known genetic associations, distortions in frequencies of microsatellite markers can focus the search for the associated gene or group of genes. It was estimated that for studies like this on families with multiple affected siblings DNA samples from at least 200 sibling pairs should be obtained and screened in order to get significant associations. These were the reasons why the study on chronic HBV carriers was initiated in The Gambia. My main task was to recruit sufficient number of sibling pairs for the human genetic study.

After associations are found by genome wide scan the candidate genes can be chosen from disease associated area of the genome. Later, in disease association studies, a candidate gene polymorphism is studied in a group of individuals exposed to a particular infection. The frequencies of the polymorphism are then compared between those with or without outcome. Very large patient groups have to be recruited for this type of case- control studies. For example, for an allele with a frequency of 10%, 307 case-control pairs are needed to have an 80% chance of detecting a twofold disease association (Thio L, 2000).

2.6.4 HBsAg carriage in The Gambia

2.6.4.1 *Studies in The Gambia*

I shall review in more detail the studies that have been conducted in The Gambia, because the information from databases of these studies was used to find chronic HBV carriers, already known to be residing in The Gambia.

2.6.4.2 *Chronic carriers in Keneba and Manduar*

These two villages, Manduar and Keneba, are situated 8 km apart in The West Kiang District of The Gambia along the river Gambia. Accurate demographic data for these two villages have been kept since 1951 (Billewicz et al, 1981). The ages were known for individuals born later than 1951 and annual surveys have revealed the seasonal pattern and prevalence of infectious disease. Between March, 1972, and March, 1973, sera from 1317 people of these two villages were tested for HBsAg. 174 (13%) were positive with the majority of cases in the 10-14 years group (McGregor et al, 1976).

Year 1980

All mothers and children under the age of 15 were asked to participate in the study. Attendance ranged from 73% for those aged 10-14 to 100% for those aged 5-9. 63 out of 174 HBsAg positive people in 1972 were tested again.

The prevalence of HBV infection was found to be markedly different in the two villages. 62 % of children in Manduar aged 2-4 years were infected whereas in Keneba only 27% of this age group were infected- none under 6 months of age and only 2 of 58 between the

ages of 6 and 12 months. The carriage of HBsAg was high, reaching a peak of 36% in the 5-9 age group in Manduar and 17,8% in the 2-4 age group in Keneba. Infection clustered in families, transmission from sibling to sibling was suspected to be of major importance. From 4 HBeAg positive mothers 10 out of 11 children became HBsAg carriers. 63% of those carrying HBsAg in 1972 were still positive in late 1980.

Age (Years)	Manduar			Keneba		
	HBsAg	%	Proportion	HBsAg	%	Proportion
	+ve		HBeAg +ve	+ve		HBeAg +ve
0-1	5/39	12,8	3/3	4/85	4,7	4/4
2-4	15/60	25,0	9/11	18/102	17,6	9/11
5-9	26/72	36,0	14/21	18/164	10,9	8/12
10-14	9/26	35,0	0/5	12/72	16,4	3/12
Mothers	20/58	34,5	4/20	11/111	9,9	0/11
	75/255		30/60	63/534		24/50

Table 3. Prevalence of HBsAg and HBeAg in two Gambian villages (All HBsAg positive samples were not tested for HBeAg).

(Whittle et al, 1983)

The prevalence of HBsAg and HBeAg carriage in persons bled during the survey in Keneba and Manduar is shown in table 3.

Year 1984

All children in the villages <20 years of age were asked to participate in the study. 306 children in Manduar and 602 children in Keneba were tested. 172 (85%) of 207 Manduar children and 326 (77%) of 424 Keneba children who were tested in 1980 were tested again in 1984.

Number of HBsAg positive children per family	Number of families in Keneba(n=142)	Number of families in Manduar(n=108)
0	88	50
1	35	32
2	13	11
3	6	9
4		6
	142	108

Table 4. Family clustering of HBsAg carriage in Keneba and Manduar

(Whittle et al, 1990)

The rate of HBV infection in the 2-4 years age group differed markedly again: in Manduar, 68%(46/68) of children became infected in contrast to Keneba 29 %(41/140). Male children were more frequent carriers of either HBs or HBeAg than were female children. Significant clustering of hepatitis B surface antigenemia within siblings was shown in both villages ($p < 0,001$). This clustering remained apparent when only HBV-infected children were considered ($p < 0,01$). Regardless of HBsAg status, HBV infection was found to

cluster in Keneba but not in Manduar. The chance of the youngest child in a household being a carrier of HBsAg was strongly related to the number of antigen positive siblings. Four years later, 53% of children who were initially positive for HBsAg and 33% who were positive for HBeAg still carried these antigens. All non immune children under the age of 5 years were vaccinated against HBV.

Year 1989

In May 1989, blood was also taken from individuals less than 20 years old and their mothers, who had not been previously tested for HBV infection (Whittle et al, 1991). 94% of those between the ages 0-4 years were tested compared with 85% of those aged between 15 and 19 years. Out of 433 vaccinees who were vaccinated with 4 different vaccination regimens since 1984 358 were traced and bled.

This study has shown that the dose, route and schedule of vaccination were of little importance in protecting children from persistent infection. Only 1 vaccinee became a carrier, despite variation of peak antibody concentration from zero to over 10 000 mIU/ml. 9% (32 out of 358) were found to be positive for anti-HBc antibodies and were thus defined as having a breakthrough infection. The number of infections after vaccination varied according to endemicity in the two villages: Manduar, which has twice the HBV infection rate of Keneba, had a fourfold higher rate of breakthrough infections. Vaccinated children of HBsAg positive mothers were at increased risk of infection, but this was not a case for children whose older siblings were HBsAg positive. The study also showed that the mother's HBsAg status was contributing to family clustering of HBV infection, probably by increased circulation of the virus in these families.

The decline of anti-HBs following vaccination was rapid, and the rate of decline was constant and independent of peak antibody concentration.

Year 1991

In May 1991, 32 children with breakthrough infections in 1989 were bled and serological tests were performed (Fortuin et al, 1994). 2 of them were still HBsAg positive. 15 children out of 32 lost their anti-HBc antibodies. 7 out of these 15 in 1989 were 1-2 years old. Three samples were PCR positive. One virus had a mutation at residue 141(lysine to glutaminic acid) of the S gene.

Year 1993

Again all children, younger than 20 years, were bled (Whittle et al, 1995). Despite a rapid decline in anti HBs antibody titres the prevalence of breakthrough infections did not increase 9 years after vaccination. Between 1984 and 1993, vaccination reduced the prevalence of HBV infection in 1-14 year old children from 50% to 9% in Keneba and from 80% to 9% in Manduar. In 1993, there were only 3 chronic carriers among vaccinated children in Keneba and 2 in Manduar compared with expected numbers (from 1984 data) of 42 and 52, respectively. The prevalence in unvaccinated groups over the age of 14 years did not change significantly in either village.

186 chronic carriers aged 4-60 years in Keneba and Manduar and 75 younger carriers, aged 3-4, from other villages were tested for HBeAg and HBV DNA by dot blot hybridisation using phosphorus-32 as label. The prevalence of these markers of infectivity declined with age.

Age	HBeAg +ve (%)	HBV DNA+ve (%)	Total number of carriers
0-4	70	95	75
5-9	40	45	17
10-14	38	45	50
15-24	22	26	52
25-34	22	16	25
35-44	15	10	22
>44			20

Table 5. Prevalence of HBeAg and HBV DNA in chronic HBsAg carriers according to their age. (Whittle et al, 1995)

I followed up 184 chronic HBV carriers identified in above-mentioned studies in Manduar and Keneba.

2.6.4.3 Studies on risk factors for transmission

In 1988 risk factors for hepatitis B virus transmission were examined in 973 Gambian children aged 6 months to 5 years in seven neighbouring villages: Kolior, Massembeh, Genieri, Kaiaf, Jiffin, Toniataba, and Sikunda (Vall Mayans et al, 1990). A total of 867 adults were tested. Villagers lived in well-defined compounds consisting of a variable number of households. In each compound the mean number of households (1.9) and people (20.5) were recorded.

33% of children were found to be infected. 4% were acute, 10% were HBsAg carriers, and 18% had evidence of infection. Of 140 HBsAg positive children 103 (73%) were HBeAg positive, and 71% were positive 6 months later.

In sera from parents, 90% (532/592) of mothers and 93% of fathers (181/192) were anti-HBc positive. Carriers more commonly had HBsAg positive mothers (25% (18/73)) than non-carriers (12 (20/167)). Significant clustering of HBsAg carriage ($p < 0.01$) was found both within compounds and within households). There was no association of malaria parasitaemia and any hepatitis marker. A significant association was found between infection and tropical ulcer scars, and between e antigenemia and the presence of bedbugs in each child's bed. There was no association between infection and traditional scarring, circumcision, or injections.

Between 1990-92 an intervention study was done over two years in the same villages to determine the contribution of bedbugs to hepatitis B transmission. In addition fortnightly questionnaires were completed for each child to assess other possible routes of transmission.

Children found to be uninfected (hepatitis B core antibody negative for children aged 12 months or older and hepatitis B surface antigen negative for those younger) were randomised into an intervention or control group. Randomisation was by compound. Intervention compounds were sprayed with insecticide.

There were 641 children in the survey who were uninfected (320 to sprayed and 321 to unsprayed compounds). Infection with hepatitis B occurred in 30% of the children over the 2 years. The intervention, insecticide spraying of the children's dwellings, was highly effective in reducing exposure to bedbugs but there was no effect on hepatitis B infection. No other risk factor for transmission was identified despite consistent village to village variation in the rate of childhood transmission.

I followed up 43 of carriers identified during this study.

2.6.4.4 Vaccination follow up studies

2.6.4.4.1 Study in Keneba and Manduar

The Gambia Hepatitis Intervention Study was a large-scale intervention study to examine the efficacy of vaccination against HBV in preventing chronic carriage of the virus, chronic liver disease and liver cancer.

In 1984, children under the age of 5 years in Keneba and Manduar villages, who possessed no serum HBV core antibody, were randomly allocated to one of three regimens of vaccination (Whittle et al, 1987).

Three trials of intradermal HBV vaccine were carried out in Gambian children. In the first trial HBV vaccine (1µg) was given to neonates in the same syringe with BCG, followed by

two further doses of the vaccine. The trial was a failure, since 19 of 32 subjects had HBV surface antibody response less than 10 m.i.u./ml.

In the second trial in young children two different regimes were used: either two doses of 2µg HBV vaccine given intradermally after a 20µg intramuscular dose or three doses of 2µg given intradermally. In both cases geometric mean antibody titres were significantly lower than in the control group who were given 20µg HBV intramuscularly followed by two 10µg doses intramuscularly. Vaccine failures, defined as presence of HBV surface antigen or core antibody or absence of surface antibody, were also significantly higher in the intradermal groups.

In the third trial 4µg of vaccine were given intradermally to 20 young children with a multiple orifice head fired by a jet gun: all had a good HBV surface antibody response > 100 m.i.u/ml in serum.

2.6.4.4.2 The Gambia Hepatitis Intervention study (GHIS)

In 1986 a national hepatitis B vaccination program of young infants was initiated gradually over a 4-year period. The country was stratified into four geographical zones of approximately equal populations. Each new team beginning vaccination was selected randomly from teams not yet vaccinating against HBV, until national coverage was achieved in February 1990. The randomisation was stratified by zone to ensure that the introduction was balanced geographically. All children attending for vaccination were registered for this period, providing two cohorts of about 60,000 children each; one

received routine vaccinations and the other received HBV vaccination too (Anonymous, 1987).

1990. To assess the response to the vaccine and the durability of vaccine-induced immunity, 1041 children from the cohort that received the HBV vaccine were followed up annually for 4 years (Fortuin, 1993). Blood samples for the 4 years of follow-up were available for 720 of these children. 33(4,6%) showed evidence of infection, 4(0,6%) of these had become chronic carriers. 9 (31%) of the mothers of children with uncomplicated infections were HBsAg positive and 5(17%) were also HBeAg positive. Only 78 (11%) of 687 mothers of uninfected children were HBsAg positive and 6(1%) were also HBeAg positive. Of 4 children who were chronic carriers, 3 had HBsAg positive mothers, 2 of who were also HBeAg positive. Overall, vaccine was estimated to be 84% effective against infection and 94% effective against chronic carriage.

To assess vaccine efficacy, a single cross-sectional study was undertaken from September, 1990, to July, 1991, in 3-4 year old children in the cohort that did not receive the vaccine (Fortuin et al, 1993). Twenty clusters, each containing 10 children, were selected from an area within each of the four geographical zones. 816 children were bled. 71% were uninfected, 29% had anti HBc, and 103 (13%) were also positive for HBsAg. Those HBsAg positive were traced a year later and 13 (14%) had become HBsAg negative.

Another cross-sectional study was carried out from September 1995 to September 1996 on unvaccinated children to determine vaccine efficacy against HBV chronic carriage and infection (Viviani et al, 1999). At nine years of age, 675 (65%) children originally recruited were traced and bled. In children 9 years of age, eight percent of the vaccinated

children had primary HBV infection compared to 50% of the unvaccinated control group; HBV carrier status was 0,6% and 10% respectively, resulting in vaccine efficacy of 83% against infection and 95% against chronic carriage. Among the vaccinated children, induced surface antibody had declined to geometric mean concentration of 19 IU/l, with 32% of the vaccinees now having undetectable antibody levels.

159 of the carriers identified in 1990 and 117 of the carriers identified in 1995 were followed up by me and used as index cases.

2.6.4.5 E-study

The study was designed to investigate factors which determine prolonged HBe antigenemia and also effects of prolonged antigenaemia (Hall, personal communication). Using data from previous GHIS studies 38 HBeAg positive women of childbearing age were selected, as well as the same number of HBeAg negative chronic HBsAg carriers and non-carriers. Subsequently, all family members of these women who could be found were included in the study.

The prevalence of HBV infection was found to be higher in mothers of HBeAg positive women (17%) than in mothers of HBeAg negative (9%) and non-carrier woman (4%), but this difference was not statistically significant. None of the mothers was HBeAg positive. In the siblings of the index woman, the prevalence was similar. The prevalence of HBsAg was significantly higher in siblings of HBeAg positive (34%) and HBeAg negative (35%) than siblings of non-carriers (13%). In fathers and husbands, no difference was found. The number of siblings in each family had an effect on the risk of the HBsAg carriage. For every extra sibling, the risk of the carrier state increased by 40%. The risk for index

women being an HBeAg positive or HBeAg negative carrier was higher in first born women and decreased with increasing birth order rank.

The prevalence of HBV carriage was significantly higher in children of HBeAg positive women (54%) than in HBeAg negative (26%) and non-carriers (19%).

26 of the carriers identified during this study were followed up by me and used as index cases.

2.7 Molecular epidemiology of the intrafamilial transmission and clustering of HBV infection.

So far, only a few studies have addressed the question of intrafamilial transmission and clustering of HBV by sequence analysis or DNA hybridisation with synthetic oligonucleotides. Neither classical subtyping nor genotyping is applicable to epidemiological tracing of transmission because of their dominance in particular communities. Finer divisions can be achieved by sequence analysis constructing phylogenetic trees or looking for distribution of specific point mutation within the family.

Two Chinese families carrying the same point mutant from C to T at position 2735 for three generations, have been described (Lin et al, 1990) and this observation was at that time considered as direct proof of intrafamilial spread of this virus.

A different selection of precore defective mutant virus between members of a Turkish family has been described (Barlet et al, 1994). The predominant strain in the mother and one of the children was the precore mutant type. In the other three children, the wild-type virus represented the major or exclusive population. The presence of an X gene deletion in the mother and three of her children was also detected.

In Taiwan, a family cluster carried immune escape variant of HBV infecting the mother and her two fully immunised children (Ho et al, 1995).

In another study, 59 Chinese families with at least two PCR positive family members showed that most of the same family members shared the same precore mutations (Akarca et al, 1994).

Extensive intrafamilial clustering of HBV variants with mutations in the pre-C and pre-S regions and long term persistence of some mutants as stable strains (Santantonio et al, 1997) have also been reported. The ethnic origin of these families was not stated.

Sequence analysis of HBV surface and core genes of 34 DNA positive members of 22 Asian new immigrant families in Australia described evidence of mainly vertical transmission and a few cases of possible horizontal transmission (McIntosh et al, 1998). Distal X and pre-Core sequence data provided better discrimination between different family groups than S gene sequences.

A study of 42 HBsAg positive subjects from 11 Chinese families, where perinatal transmission was thought to be predominant, showed a higher degree of homology within families than between families. Within all families, core sequence homology was higher than 97.5%. In two families, all family members had completely identical sequences over the 20-year period after infection. Two identical pairs were found in other families as well. In other families, the increased divergence of the sequences was associated with age and HBeAg status.

A study of 19 Italian families showed that most of the infected family members shared the same virus strain with very high homology (Karayiannis, personal communication). For

detection of viral strains infecting each particular family sequencing of the pre-core/core genes was found to be the most informative.

2.8 *Phylogenetic methods*

Phylogeny studies convert information from nucleotide and amino acid sequences into an evolutionary tree for those sequences. Several methods have been developed and are classified into distance and discrete methods. The division is based on how the data are treated; distance methods first convert aligned sequences into a pair wise distance matrix, then input the matrix into a tree building method, whereas discrete methods consider each nucleotide site directly.

The most commonly used distance methods are the neighbour joining (NJ) and unweighted pair group method with arithmetic means (UPGMA). The NJ method is more commonly used, because it is very fast and is considered the best heuristic method for a small number of substitutions but it does not differ substantially from more conventional methods. The NJ tree can give sufficient insight into what the final tree will look like and help in deciding which more advanced methods should be used.

The main discrete methods are maximum likelihood (ML) and maximum parsimony. ML is considered by many to be the current method of choice for tree reconstruction (Clewley, 1998).

The phylogenetic analysis of molecular sequence data using explicit evolutionary models and the maximum – likelihood approach has been used since 1981 (Felsenstein et al,

1981). This approach is now widely used and has seen many applications, from looking at viral divergence within a single patient (Holmes et al; 1992) to examining the relationships between mammalian orders (Novacek et al, 1992). The maximum likelihood approach has not been used for evolutionary analysis of HBV until very recently (Bollyky et al, 1999).

The tree that makes the most probable evolutionary outcome is the maximum likelihood estimate of the phylogeny. Likelihood says nothing about the probability of the model itself. Given a model that specifies probabilities of observing various events, the likelihood can be computed. This can be written as:

$$L = \Pr(D/H)$$

where $\Pr(D/H)$ is the probability of getting the data D given hypothesis H. As L is often a very small number, likelihoods are often expressed as natural logarithms and referred as log-likelihoods. This mathematical expression makes likelihood a powerful tool for hypothesis testing. Also it provides a mathematical tool to see which of the chosen nucleotide substitution models results in a greater increase in likelihood (Huelsenbeck, 1997).

Given that observed distances may underestimate the actual amount of evolutionary change, there has been a considerable amount of research in developing methods of converting observed distances into measures of actual evolutionary distance. These techniques are termed distance correction methods, which use different substitution models. Transition/transversion bias and unequal base frequencies have to be taken into account in order to obtain higher likelihoods. The relative degree of fit between each model and the data can be evaluated by computing for each model (see table 6) the

likelihood of obtaining the observed data. For HBV infection, the HKY 85 substitution model seems to be the most appropriate and commonly used (Bollyky et al, 1999).

Substitution model	Base frequencies	Transversion/transition ratio
Jukes-Cantor (JC)	Equal	All substitutions equally likely
Kimura 2 (K2P)	Equal	Transversions and transitions have different substitution rates
Felsenstein (F81)	Unequal	All substitutions equally likely
Hasegawa et al (HKY85)	Unequal	Transversions and transitions have different substitution rates
General reversible (REV)	Unequal	All six pairs of substitutions have different rates

Table 6. Interrelationships among the five best-known models for estimating the number of nucleotide substitutions among a pair of DNA sequences.

Different sites have a range of probabilities of substitution. It was difficult to develop tractable models of this rate variation. The most widely used approach makes use of the gamma distribution. This distribution has 'shape' parameter α , which specifies the range of rate variation among sites. Small values of α result in an L-shaped distribution with extreme variation of rates. Conversely, the larger the smaller the range of rates. These α values on different genes of HBV are relatively small reflecting quite different rates of substitution in different parts of the genome (Bollykyn et al, 1999).

One reason for a poor estimate of a phylogeny may not be the method used, but limitations in the data themselves. Estimates of phylogeny based on samples will always be accompanied by sampling error. Trees are complicated structures and it is extremely difficult to develop equations for confidence intervals around a phylogeny. An approach called bootstrap mimics a multiple sample approach with resampling from the sample. Each resampling is a pseudoreplicate. Bootstrapping can be applied to phylogenies by generating pseudoreplicates from the sequence data. Pseudoreplicates will resemble the original data set in that they contain only sites found in the data set, but will differ in frequencies at different sites. From these pseudoreplicates it is possible to build trees using conventional methods. The process can be repeated in a large number of times (up to 1000 fold) resulting in a set of bootstrap trees. This set of trees contains information on the sampling error associated with our sample. The most common splits found among the bootstrap tree can be assembled into a bootstrap consensus tree. These are often drawn with each node labelled with its frequency of occurrence among the bootstrap trees. Originally it was suggested that only nodes with bootstrap values above 95% should be

accepted and well supported (Sanderson, 1995), but currently values higher than 70% are considered as having a statistical significance (Holmes E, personal communication).

Maximum likelihood provides a natural means for hypothesis testing. Likelihood ratio tests (LRTs) are known to be optimal when comparing simple hypotheses and the same time they often perform well for cases in which no optimal test is known. The likelihood ratio provides a measure of the support of the data for one hypothesis versus another.

Null distributions can be generated by using computer simulation (Felsenstein, 1988). Parametric bootstrapping or Monte Carlo simulation of phylogenetic trees has been previously described (Huelsenbeck, 1997). Program package PAUP 4.0 has a tree simulation option and generation of simulated trees from the given phylogenetic tree is easy to perform. ML logarithmic values from each simulated tree can be obtained and then compared to other logarithmic values obtained from the real ML trees.

Explicit model-based methods are a recent innovation in phylogenetics and published information is very limited, but it is clear that these approaches allow a much broader evolutionary analysis.

2.9 Geographical setting of the study

2.9.1 The country

The Gambia lies on the west coast of Africa, between latitudes 13 ° W and 17 ° W. The country forms a narrow band on either side of the River Gambia and is almost entirely surrounded by the Republic of Senegal (see figure3). The Gambia averages about 24 km in width and extends along both banks for about 487 km. The total area of The Gambia is 10367 km².

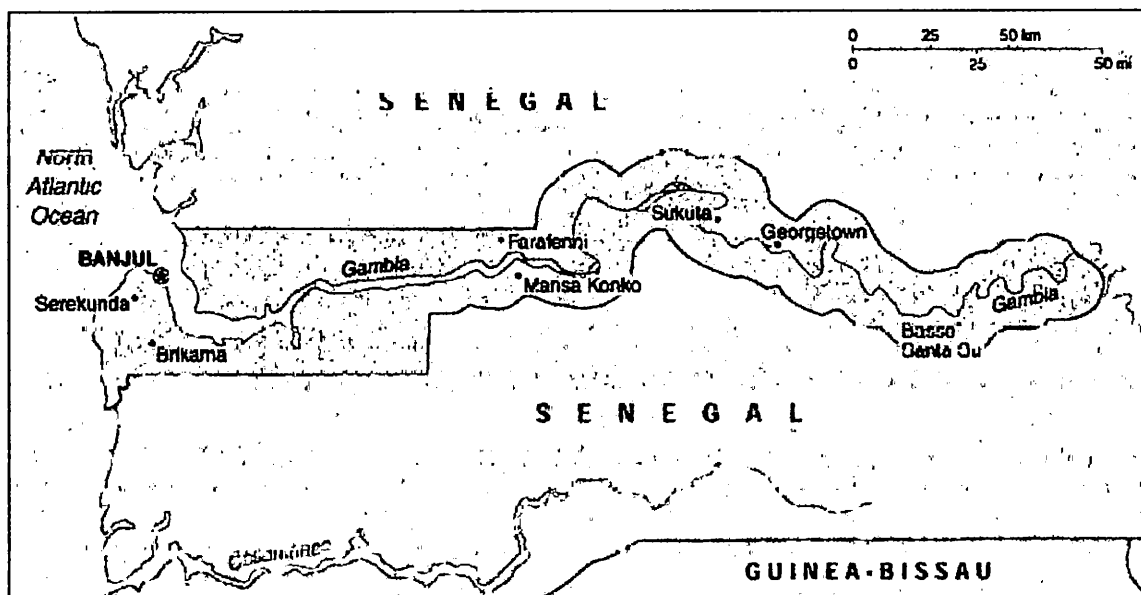


Figure 3. Map of the Gambia

The Gambia a population of is 1038145 inhabitants according to the 1993 Census. The highest concentration of the population is in the periurban areas of Banjul, Serekunda and Brikama. The main ethnic groups are Mandinka (40%), Fula (19%), Wolof (15%), and Jola (10%). Other ethnic groups as Manjago, Serere, Sarahule are in minority. The annual

population growth rate is 4,1% and the population density is 46 people per square kilometre. The vast majority of Gambians are Muslim.

Administratively, The Gambia is divided into 5 Divisions and 2 Municipal Areas. The two municipal areas are Banjul and Kanifing, each including the periurban areas around the coast. The divisions are the Western Division, the Lower River Division, the Central River Division, the Upper River Division and the North Bank Division.

The climate of the Gambia is tropical with two distinct seasons: a dry season from November to June, and the wet season from June to October.

2.9.2 Geographical distribution of villages involved in study

The study referred to in this thesis was carried out throughout all Gambia. The main MRC laboratories in Fajara and Basse, and the Farafenni field stations were used to access known carriers from each village.

2.9.3 Ethnic groups participating in the study

The main ethnic groups represented in the study are Mandinka Wollof, Fula, Jola, Sarahulé and Serere. Some refugees from countries such as Sierra Leone, Guinea Bissau were occasionally included in the study.

2.10 Study Objectives

To characterise the family clustering of HBV carriage in Gambian families by analysing the prevalence of HBV serological markers in siblings according to their age, birth order, sex, and serological status of their parents.

To characterise predominant genotypes and subtypes of HBV in The Gambia by obtaining consensus sequences of HBV surface, core promoter, and pre-core/core genes and confirm or exclude the presence of HBV mutations with clinical significance.

To employ phylogenetic and statistical methods of HBV sequence analysis to confirm hypothesis on horizontal transmission of HBV within Gambian families by combining it with epidemiological data.

3 Methods

3.1 Field work

The field work was done with the help of two field workers- Mr.Lamin Giana and Mr. Joseph Bass. Patients were usually visited three times by the field worker and once by me for the bleeding.

3.1.1 Finding of carriers, getting consent from family members and making appointments

Data about carriers were taken from the databases of previous studies in The Gambia. Information about the carrier's name, surname, name of their parents and village of residence were usually available. The field worker had to find the carrier and talk to his/her parents. All siblings of the index case older than 10 years (not vaccinated against Hepatitis B) and parents were asked to take part in the study and give a blood sample. The consent form was completed for every blood sample taken.

In some families all family members were bled during the previous surveys (Keneba and Manduar), so siblings carriers were already identified. On such occasions only those known to be positive and their parents were bled.

Following verbal explanation the consent form was signed by the field worker on behalf of every person participating in the study.

Example of the consent form shown as Appendix I

At the same time, a questionnaire was completed to document the family status, how many siblings were vaccinated, how many were not present and where they could be found.

An example of the family field questionnaire shown as Appendix II.

After the first visit, consent was obtained and appointments were made for bleeding on the next occasion.

3.1.2 Bleeding

5-10 ml of blood usually was taken through venepuncture from each patient. 1ml of the specimen was stored in a serum separator (Becton Dickinson) and the rest of the blood in a 10ml EDTA tube. Specimens were stored in a cool box until arrival to the laboratory.

Before bleeding the questionnaire was completed for every person bled. The field worker completed the left side of the questionnaire and I filled the right side later in order to check for mistakes. Example of field questionnaire shown as Appendix III

Immediately after arrival from the field, both tubes were spun at 3500 rev/min for 10 min. Serum in microseparators was stored at -20°C . Plasma from EDTA tubes was immediately separated and stored at -70°C .

3.1.3 Delivery of preliminary results

Serum was tested for HBsAg status and the field worker delivered the form shown in Appendix IV to parents or to every family member separately. He explained the

consequences of a positive result and also informed them why they will be visited later again for follow up studies.

3.1.4 Data entry and statistical analysis of epidemiological data and laboratory results

All information from the field questionnaires and laboratory sheets were entered into EPI INFO version 6.02 program package. The same package was later used for statistical analysis.

3.2 Laboratory work

3.2.1 Hepatitis B serology

3.2.1.1 HBsAg detection

Reverse passive agglutination (Hepato-test, Wellcome Diagnostics) was used to detect HBV surface antigen. Positive tests were confirmed by neutralisation reaction, according to the manufacturer's instructions.

Despite RPA being less sensitive than other conventional methods (RIA, ELISA) it was decided, that the shortfall in study like this would not be substantial in achieving the main aim; collection of families with multiple affected siblings. The other consideration was that we had to bring preliminary results as soon as possible back to the families bled. The RPA for field conditions was much more convenient and the number of samples tested can be very small. Previous studies in The Gambia showed that the difference in sensitivity between RPA and RIA is around 3-5% (Whittle H, personal communication).

All negative samples were later retested for hepatitis B surface antigen with an immunoradiometric assay (Sorin Biomedica) to achieve highest possible sensitivity.

3.2.1.2 Other HBV serology

An HBeAg/anti-Hbe immunoradiometric assay kit (DiaSorin) was used according to the manufacturers instructions for detection of HBeAg, anti-HBe, Total anti-Core, and anti-Core IgM.

3.2.2 PCR and sequencing

3.2.2.1 Buffers and solutions

Ethidium bromide	10mg/ml ethidium bromide solution in sterile deionised distilled water
1% Agarose gel solution	1g of Agarose to 100 ml TBE
TBE buffer	0.9 M Tris, 0.9 M Boric Acid, 25mM disodium EDTA (pH=8,3)
dNDPs	20 µl for each dNTP (100mM solution) 320 µl of water
Chloroform	Chloroform/isoamyl alcohol (25:1) equilibrated pH 8
Phenol	Saturated with 0.1M citrate buffer pH 4.3, equilibrated with water (pH adjusted to 7.5)
Loading buffer for automated sequencing	99.5% deionized formamide and 25 mM EDTA containing 50mg/l dextran blue in a ratio of 5:1 formamide to EDTA/ dextran blue
2.5X sequencing buffer	200mM Tris/HCl pH=9, 5mM MgCl
6X agarose sample loading buffer	0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% sucrose in deionised distilled water

Table 7. Buffers and solutions

Films and photographic reagents

Kodak instant films were used for photographing of gels.

Molecular weight markers

100bp DNA ladder consisting from 11 double-stranded fragments with sizes of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500 (Promega) was used in all agarose gels as a molecular weight marker.

3.2.2.2 DNA extraction

	HBeAg positive	HBeAg negative	Stock solution
Serum or Plasma	50 μ l	150 μ l	
10X buffer	20 μ l	40 μ l	250mM NaCl/25mM EDTA
Proteinase K	20 μ l	40 μ l	20 mg/ ml
t-RNA	2,5 μ l	5 μ l	8 mg/ ml
10% SDS	20 μ l	40 μ l	10 g/ 100ml (10%)
Water	87,5 μ l	175 μ l	

Table 8. DNA extraction mix

The mix was incubated at 37° C overnight, extracted twice with Phenol/Chloroform and once with chloroform, before precipitation of the DNA by adding 1/10 of the volume 3 M Na Acetate and 2 volumes of 100% ethanol at -70° C for one hour. Following centrifugation the pellets were washed with 70% ethanol and dried before resuspending the pallet in 20µl of distilled water.

3.2.2.3 PCR amplification of HBV DNA

3.2.2.3.1 PCR reaction

The following reagents were mixed in eppendorf tube for the amplification of the target region by PCR.

		Stock	Final concentration
DNA	5µl		
Buffer	10µl	10X PCR buffer	100mM Tris-HCl 500mM KCl 0,8% Nanidet P40
MgCl	6µl	25 mM	1,5 mM
dNTP	4µl	100 mM	10 mM (mix)
Primer 1	1µl		0.2 pmol
Primer 2	1µl		0.2 pmol
Water	73µl		
Taq	0,2µl	5 U/ ml	
Total	100,2µl		

Table 9. PCR reaction mix.

The thermo- cycling conditions for both the first and second round reactions

Time	Temperature	Number of Cycles
4 min	95°C	1
1min	95°C	40
1min	53°C	
2,5 min	72°C	
10min	72°C	1

Table 10. Thermo-cycling parameters of PCR

Negative samples after first PCR round were subjected to heminested PCR using another set of primers.

3.2.2.3.2 PCR of the pre-core/core gene

Oligonucleotides

M3

5'-CTGGGAGGAGTTGGGGGA Position 1732-1755 (sense)

6C

5'-GGCAAGCCATTCTTTGCTGGGG Position 2070-2092 (sense)

5C

5'-CCCACCTTATGAGTCCAAGG Position 2466-2486 (anti-sense)

Pol8

5'- AGGATAGAATCTAGCAGGC Position 2649-2667 (anti-sense)

8C

5'-GTCCCTGGATGCTGGATCTTGCT Position 2130-2154 (anti-sense)

PCR reactions

- a) M3-POL8 for the first round

Length of product, 935 bp

- b) M3- 5C for the second round

Length of product, 754 bp

Other combination of primers were tried later to improve the sensitivity of the reaction

- a) M3-5C for the first round

Length of product, 754 bp

- b) M3-8C and 6C-5C for the second round (heminested PCR)

Length of product 420bp with primers M3-8C

Length of product 416bp with primers 6C-5C

2 µl of the first round PCR product were used as template for the second round PCR.

3.2.2.3.3 Surface gene

Oligonucleotides

S3

5' CAAGGTATGTTGCCCCGTTTG Position 457-476 (sense)

S4

5' GGGTTTAAATGTATACCCAGAGAC Position 817-838 (antisense)

Reaction S3-S4

Length of product, 381 bp

3.2.2.3.4 Analysis of PCR products

PCR products were analyzed by agarose gel electrophoresis in TBE buffer. 7 µl of the PCR sample was run on a 1% agarose gel by adding 2 µl of 6X loading buffer. 4µl of 100bp DNA ladder was also added to each gel run. The gel was stained by shaking in a solution of 10 µg/ml ethidium bromide and bands were visualized under UV illumination and photographed.

3.2.2.4 DNA purification

A QIAquick purification kit was used to extract the DNA fragment from the 1% agarose gels, after running 50µl of the PCR product. To the excised gel slice 400µl of Buffer QG were added and incubated for 10 min at +50°C, the mix was poured into spin column and centrifuged for 1min, and the immobilized DNA was washed with 500µl of QG buffer and subsequently with Buffer PE, before centrifuging again. Finally, 25µl of elution buffer was added and the DNA was stored at -20° C.

After purification 5µl of the product were run in 1 % agarose gel to assess the quality and size of the purified product.

3.2.2.5 Sequencing

Automatic sequencing was done on an ABI PRISM 377 DNA sequencer (Perkin-Elmer) and sequences read with supporting software for line extraction and analysis.

Sequencing reaction

Purified PCR product	2 µl
2,5X Sequencing Buffer	4 µl
Big Dye Terminator	4 µl
Primer (1.8pmol solution)	2 µl
Distilled water	8 µl
Total	20µl

Thermo cycling parameters of sequencing reaction

1min	96°C	
0.5min	50°C	25cycles
3min	60°C	

Extension 4 °C

Sequencing gel

Acrylamide gel:

18 g	Urea
5,3 ml	Acrylamide solution
5 ml	10X TBE buffer
250µl	10% Ammonium Persulphate
35µl	TEMED
27,5ml	Water

After thermocycling, the products were precipitated by adding 50µl of 95% Ethanol and 2 µl 3M Na Acetate solution. This was allowed to stand for 10 min and then centrifuged. The pellets were washed with 70% Ethanol, centrifuged again, vacuum dried and then stored at -20°C .

Before sequencing 6 µl of sequencing loading buffer were added to the pellet and the DNA denatured at 95°C for 2 minutes. 2µl of this solution were loaded onto the sequencing gel.

3.2.3 Analysis and editing of the sequences

Editing and translating sequences to protein sequences- programme package DNASIS

Sequences obtained by automatic sequencing were edited by comparing with the published *ayw* subtype sequence (Norder et al, 1993).

Afterwards sequences were exported to Notepad format document and aligned in FASTA format for further phylogenetic analysis.

Protein sequences were translated from nucleotide sequences using DNASIS programme.

Editing and comparing protein sequences

Programme package PROSIS was used for editing and alignment of amino acid sequences.

3.2.4 Phylogenetic analysis

Sequence alignment

All sequences were aligned in FASTA format (see appendix 6). Afterwards they were changed to NEXUS format (see appendix 7) by using programme **SeqVerter** (downloaded from www.genestudio.com/seqverter.html). This alignment can be afterwards used by the **PAUP 4.02b** programme package.

Pictures of alignments used in the thesis were produced using **Multalin** program Web interface (<http://www.toulouse.inra.fr/multalin.html>) (Corpet).

Neighbour joining trees.

NJ trees on S gene sequence fragments and Precore/core gene sequences were constructed by using **PAUP 4.02b** programme package.

Maximum likelihood trees.

Two genotypes A and E were analysed separately in order to reduce sample size and make mathematical calculation more efficient. Later sequences were devised according to their HBeAg status. This resulted in four different sequence data sets; (I) all available genotype A sequences ($n = 14$), (II) all available genotype E sequences ($n = 99$), (III) HBeAg negative samples from genotype E ($n = 30$), (IV) HBeAg positive samples from genotype E ($n = 53$). Note that the sum of the E antigen negative and positive samples is less than the total number available for this genotype because, in both of these cases, sequences, which were the sole representatives from a village, were also excluded. It was not possible to subdivide the genotype A sequences by E antigen status because of the small number of samples available.

Phylogenetic relationships among the HBV sequences sampled were reconstructed using a maximum likelihood (ML) method. This method incorporated the Hasegawa-Kishino-Yano (HKY85) model of nucleotide substitution with a gamma distribution of rate variation among sites. Both the transition/transversion (Ts/Tv) ratio and the α shape parameter of the gamma distribution, which determines the extent of rate variation, were estimated during tree reconstruction. To assess the robustness of this phylogeny, and particularly whether viruses were clustered according to the family and village from which they were sampled, a bootstrap resampling analysis was undertaken (1000 replications), although computational constraints meant that this analysis was based on replicate neighbour-joining trees reconstructed using the ML substitution model.

A maximum likelihood approach was also used to determine whether sequences were clustered by family and village more than might be expected by chance alone. First, for each of the four data sets, 200 random tree topologies were constructed. The log likelihood of each of these replicate trees was then estimated, again using empirical Ts/Tv and α values. As these trees represent random assortments of the data, they effectively constitute a null distribution of log likelihoods. Next three “model” trees, representing clustering at different levels of population structure, were constructed using the **TreeView** programme: (I) one in which only sequences from the same family were grouped together, with all other branches left unresolved, (the “family” tree) (II) one in which only sequences from the same village were grouped together, with all other branches left unresolved and ignoring what family they came from (the “village” tree) and (III) one in which sequences from the same family and village were grouped together (the “family+village” tree). The log likelihoods of these model trees were then estimated under the data as before. If these likelihoods fall outside the null distribution then we could conclude that the HBV sequence data shows more clustering by family and village than might be expected by chance. It should be noted that this randomisation test was highly conservative because most branches are left unresolved, particularly when the data were only partitioned by family, and therefore ignored any other phylogenetic signal in the data. All phylogenetic analyses were undertaken with **PAUP 4.02** package (Swofford).

4 Human epidemiology of HBV and characteristics of the chronic HBsAg carriers

4.1 Introduction

Hepatitis B virus (HBV) infection is endemic in the Sahel area of West Africa (Barin et al, 1981; Whittle et al, 1983), where high rates of hepatocellular carcinoma (HCC) associated with chronic carriage of the virus are reported (Bah et al, 1990). Chronic infection with HBV acquired early in life is the major risk factor for development of HCC in humans. HCC is the major cause of cancer mortality in males in the African continent (Pisani et al, 1997). Mass vaccination against HBV, introduced in The Gambia in 1986, was very effective in preventing chronic infection in children (Viviani et al, 1999; Fortuin et al, 1993) but disease burden still remains high in the adolescent and adult population.

Epidemiology of HBV infection in The Gambia was discussed in detail in the main Introduction (see 2.6.4). It was shown previously that most of the HBV infections in The Gambia are acquired in early childhood and mother to child transmission is not a predominant route of transmission, because majority of babies do not become positive until 6 months to 1 year after birth (see 2.3.1.3). Possible horizontal transmission between children has always been suggested but the exact mechanisms of the transmission were never found. It is also known that males in The Gambia are more likely to become chronic HBsAg carriers and to develop HCC than females. Significant clustering of HBsAg positivity among siblings on the village and compound level was previously described in two Gambian villages (Whittle et al, 1983 and 1990).

The main original aim of the sample collection was to investigate the genetic determinants of susceptibility to persistent infection and disease with HBV. A large number of affected

sibling pairs was recruited for this study in order to perform the whole genome scan (see 2.6.3).

4.2 Results

1180 people from different areas and ethnic groups of The Gambia were bled during the study. Samples were taken from 351 previously known chronic HBsAg carriers and in 181 cases their family members whose HBsAg status was unknown were also bled. Children younger than 9 years were not tested because they were assumed to be vaccinated.

Previous study	Number of carriers	Sex		Age group (years)		
		M	F	9-19	20-29	30+
E-study	26	13	13	9	6	11
Keneba and Manduar study (2.6.4.2)	184	88	96	79	57	48
The Hepatitis Intervention study (2.6.4.4.2)	207	91	116	139	26	42
Studies on risk factors of transmission (2.6.4.3)	45	29	16	45		
Total	462	221	241	272	89	101

Table 11. Known index carriers bled in previous studies and their sex and age distribution.

	Fathers	Mothers	Sons	Daughters
Bled	92 (51%)	107 (75%)	321 (58%)	308 (62%)
Not bled	87	74	234	190
Total	181	181	555	498

Table 12. Proportion of family members found and bled in families where only one carrier was identified previously.

29% (125/424) of the siblings not bled were under 9 years of age and therefore vaccinated and most likely protected against the HBV infection. The rest of them were not present at the time of bleeding or refused to take part in the study. Fathers who were not bled usually were travelling or refused to take part in the study.

624 samples tested were HBsAg positive and anti Core IgM negative. Two samples were found to be anti-HBc IgM positive and therefore not defined as chronic HBV carriers. 439 (93%) out of 473 HBsAg negative samples tested for total anti-Core antibodies were found to be positive. Other 82 HBsAg negative samples were not tested because the amount of the plasma sample was insufficient (200µl was needed for each serological reaction).

Ethnic group	Total bled	Prevalence in bled	Prevalence in Gambian population
Mandinka	765	65%	40%
Wollof	95	8%	15%
Jola	63	5%	10%
Fula	85	7%	19%
Sarahule	81	7%	7%
Other	93	8%	9%
Total	1180		

Table 12. Prevalence of the ethnic groups in study subjects in comparison to the overall prevalence in The Gambia.

The proportion of Mandinka ethnic group is higher because most of the persons were bled in Basse, Brikama and Keneba areas where the Mandinkas are the predominant ethnic group.

Family clustering of HBV infection

Family clustering was defined as present if at least one sibling of the known carrier or parents were found to be chronic HBV carriers.

In 181 families HBsAg status was known only in one sibling and the rest of the family members were bled and tested. In 142 (78%) of these families at least one more sibling was found to be a chronic HBsAg carrier. In 34 families family clustering was already previously confirmed and only known carriers were re-bled. In 55 families only one previously known carrier and only one sibling with an unknown HBsAg status were bled. In 27 (49%) of families that another sibling was found to be a chronic carrier.

Chronic HBsAg carriers		Siblings bled	
Number of carriers in the family	Number of families with chronic HBV carriers	Number of siblings bled in each family	Number of families
1	64	1	70
2	85	2	36
3	51	3	49
4	10	4	22
5	5	5+	22
Total	215		222

Table 14. Numbers of siblings bled, number of HBsAg carriers found and the number of families tested. In some families the only sibling bled was found to be HBsAg negative that could be the explanation why some families did not have chronic carriers.

Prevalence of HBsAg positivity in family members of previously known carriers

	Families with multiple HBsAg carriers		Families with at least one carrier	
	HBsAg positive	Tested	HBsAg positive	Tested
Fathers	8 (11%)	72	14 (12%)	114
Mothers	24 (24%)	99	33 (23%)	144

Table 15. Prevalence of HBsAg positivity in parents from all families with at least one sibling carrier and from the families with multiple siblings. (11 mothers from E study (2.6.4.5) are excluded because they were already known HBeAg carriers 6 years ago and their children were bled in order to characterise mother to child transmission).

Mothers were more likely to be the carriers than fathers ($p=0.01$, χ^2 test) (see Table 15) in the families with multiple affected siblings and in the families of single carriers ($p=0.02$, χ^2 test). The HBsAg prevalence in the fathers was similar to the rate reported in the normal population.

In this study only four mothers of the carriers were found to be HBeAg positive. That was not enough to do any statistical analysis on the influence of HBsAg status of the mother.

Siblings	HBsAg prevalence in tested	Tested	Not tested
Sisters	37%	203	204
Brothers	57%	201	248

Table 16. HBsAg status according to sex of the siblings of the chronic carrier. (Data from 21 families where HBsAg status of family members was known previously are excluded).

These data (see table 16) show clearly that chronic Gambian HBsAg carrier has a very high chance of having a sibling carrier, which is much higher than overall prevalence of the chronic carriage (10-15%) in The Gambia (Whittle, 1990).

Brothers of the carriers were statistically more likely to be carriers ($p < 0.001 \chi^2$ test) than sisters of the carriers. The relative risk of being a carrier if you were a male sibling versus a female one was 2,1 (1,4- 3,1). After adjustment for age using logistic regression approach this was completely unchanged i.e. age made no difference.

HBsAg carriage according to the HBsAg status of other family member

From the family data available it was very difficult to analyse carriage rates according to the birth order, family size and HBsAg or HBeAg status of the mother, because too many family members were not bled or were vaccinated and therefore protected against infection. However, it was possible to see the distribution pattern of the infection within the families with multiple chronic HBsAg carriers.

If we assume that transmission from mother to the child occurs, then we expect that children born first are more likely to become chronic HBsAg carriers, because at the younger age mothers are more likely to be HBeAg positive with higher viral loads. Only two mothers from families with multiple HBsAg carriers were HBeAg positive, but their status at the time of actual transmission of the infection was not known. The prevalence of HBsAg positives in the mothers of multiple carriers is significantly higher than in fathers or at the population at large. That indicates a possible link between mother's and child's HBsAg status.

Birth order	HBsAg carrier mother (N=13)	Non-carrier mother (N=44)	Mother's status unknown (N=153)
1	1	1	1
2	0.4 (0.1-2.5)	2.4 (0.8-7.0)	1.5 (0.8-2.8)
3	2.8 (0.5-17.3)	3.0 (1.0-8.5)	2.7 (1.4-5.3)
4	0.8 (0.1-7.3)	5.5 (1.7-17.9)	1.7 (0.8-3.5)
5	-	5.0 (1.3-19.5)	1.9 (0.8-4.5)
6	1.6 (0.1-38.0)	6.0 (1.2-29.2)	6.6 (1.5-27.9)
7		7.0 (1.4-35.5)	4.9 (1.2-20.1)
Test of trend	p=0.4	P=0.003	P<0.001

Table 17. Relation of birth order adjusted to family size and HBsAg status of mothers.

If we assume that children became infected from older HBeAg positive siblings then the carriage rate in younger siblings should be higher, because they are getting exposed to the virus earlier and therefore are in higher risk of becoming carriers. Many families were

clearly following this pattern where several first siblings were HBsAg negative and their younger siblings were carriers (see Appendix V, families 40, 108), but this was not enough to draw a conclusion that in the families with multiple affected siblings younger children were in higher risk of developing a chronic infection. Also, I did not know whether siblings had grown up in the same compound or had travelled during childhood.

Analysis in Table 17 shows that birth order was not a significant risk factor for the chronic HBsAg carriage in families with a carrier mother, but the difference was significant for the trend in families where mothers were non-carriers or their carrier status was unknown. The number of families was small but these data indirectly indicated that acquisition of the HBV from older siblings takes place in families with multiple chronic HBsAg carriers.

Clustering of the HBeAg in Gambian families

I defined a clustering of the HBeAg in the family as a presence of two or more HBeAg positive family members in the family. In 24 families clusters of the HBeAg positive siblings were identified. They were not always the youngest siblings. Some HBsAg carriers already seroconverted to anti-HBe but their older siblings remained HBeAg positive. For example, in family 174 only oldest and youngest out of three sisters HBsAg carriers bled were HBeAg positive (see Appendix V).

Extremely interesting were families 141, 61 and 217 (see Appendix V) where all siblings tested were found to be HBeAg carriers. These families are of special interest for future molecular epidemiology and human genetic studies.

Age (Years)	Males		Females	
	HBeAg positives (%)	Total	HBeAg positives (%)	Total
10-14	48	70/147	30	25/84
15-19	37	21/56	21	9/42
20-29	12	5/43	13	8/63
30-39	15	2/13	5	3/55
40+	10	2/20	3	1/33
Total	36	279	17	277

Table 18. Prevalence of HBeAg positives in HBsAg carriers adjusted to their age and sex. Selection of these age groups better reflects the dynamics in age group 10 to 20 where the most of the seroconversion occurs.

The sex difference in the HBeAg carriage was significant for the trend adjusted for age. Therefore boys remain as potential reservoirs of infection for longer. It is also evident that most of the patients become HBeAg negative at the age of 15-20 when most of seroconversions occur.

	Surface gene (% positives) First PCR	Core promoter/pre-core/core gene (% positives)		
		First PCR	Second PCR on 1-st negative	Total with core primers
HBeAg positives	83%(83/99)	54% (54/100)	80% (37/46)	91%(91/100)
HBeAg negatives	27% (44/160)	14% (21/155)	48% (60/124)	52%(81/155)
Total	50 % (127/259)	33% (75/255)	57% (97/170)	73%(186/255)

Table 19. PCR positivity according to HBeAg status

The PCR on pre-core/core genes was performed with different primer sets. Though sensitivity of the assay at the first PCR varied from 20 to 85 % during the optimisation of reactions the total sensitivity of both reactions was acceptable in comparison with other studies (Karayiannis, personal communication). The surface gene PCR worked better from the very beginning and hemi-nested PCR was not performed. This means that the overall sensitivity of this assay is lower than on pre-core/core reaction. PCR on pre-core/core and surface genes were not always performed on the same samples so the data shown in table 19 should be interpreted cautiously.

Age	PCR positives in males	PCR positives in females	Total prevalence
10-14	73/87 (84%)	38/51 (74%)	111/138 (80%)
15-19	34/50 (68%)	11/32 (34%)	45/82 (55%)
20-29	19/36 (53%)	20/40 (50%)	39/76 (51%)
30-39	2/9 (22%)	10/24 (42%)	12/33 (36%)
40+	4/11 (36%)	6/16 (37%)	10/27 (37%)
	139/200 (69%)	86/166 (51%)	225/366 (61%)

Table 20. Prevalence of the PCR positivity adjusted to age and sex (PCR results of surface and core genes are combined).

PCR positivity could also serve as a good marker for high viral load. Table 20 cannot give a completely objective picture, because the sensitivity of PCR assay improved significantly during the optimisation. The purpose of this study was not to get higher sensitivity, but to get sequences from affected siblings and their parents. The high prevalence of PCR positives in the group of women aged 20-39 could be biased, for 11 of them were part of the mother to child transmission study (E-study) and 6 years ago were already known to be HBeAg positives. The PCR on pre-core/core gene was started on samples from the younger age group and HBeAg positives and then after optimisation it was done on the older age groups. Therefore the sensitivity of the assay performed on older age group was higher.

The decline of the PCR positivity according to age is still obvious. There is a statistically significant difference between the rates of PCR positives in males and females in the age group from 10 to 19 years, corresponding to the sex difference in rates of HBeAg carriage.

Levels of aminotransferases

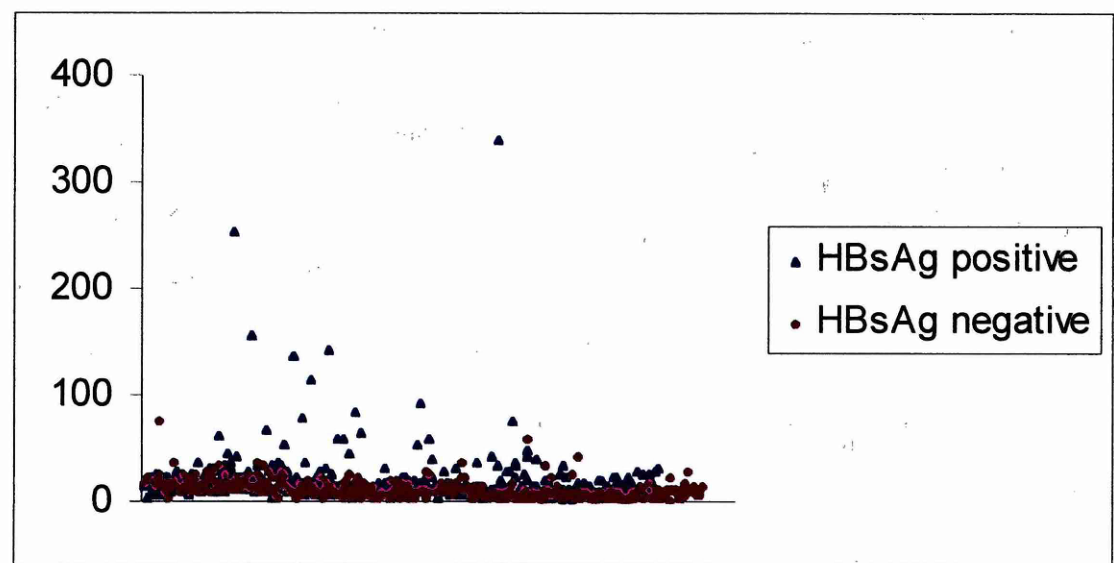


Figure 4. ALT levels in HBsAg positive and HBsAg negative patients.

ALT (IU/ml)	HBsAg negative	HBsAg positive
0-39	439	385
40+	3 (1%)	27 (7%)
Total	442	412
Mean age (Years)	30	21

Table 21. Levels of ALT in HBsAg positive and HBsAg negative persons

Internationally accepted upper limit of the assay is 40 IU/ml. The difference between a HBsAg positive group and HBeAg negative group is statistically significant ($p<0,001$, χ^2 test).

ALT	HBsAg positive		HBeAg positive		HBeAg negative		HBsAg negative	
	Males	Females	Males	Females	Males	Females	Males	Females
0-39	197	188	67	32	124	147	184	255
40+	19 (9%)	8 (4%)	11(14%)	4 (11%)	8(6%)	4 (3%)	2 (<1%)	1(<1%)
Total	212	196	78	36	132	151	186	256
Mean age	19	25	13	17	20	26	34	28

Table 22. ALT values and patients HBsAg and HBeAg status in relation to the gender of the patients.

ALT values are more often elevated in male carriers ($p=0.05$, χ^2 test). But HBsAg positive males bled were on average younger then females and probably have higher viral replication, which could lead to elevated enzyme levels. This has not been reported before.

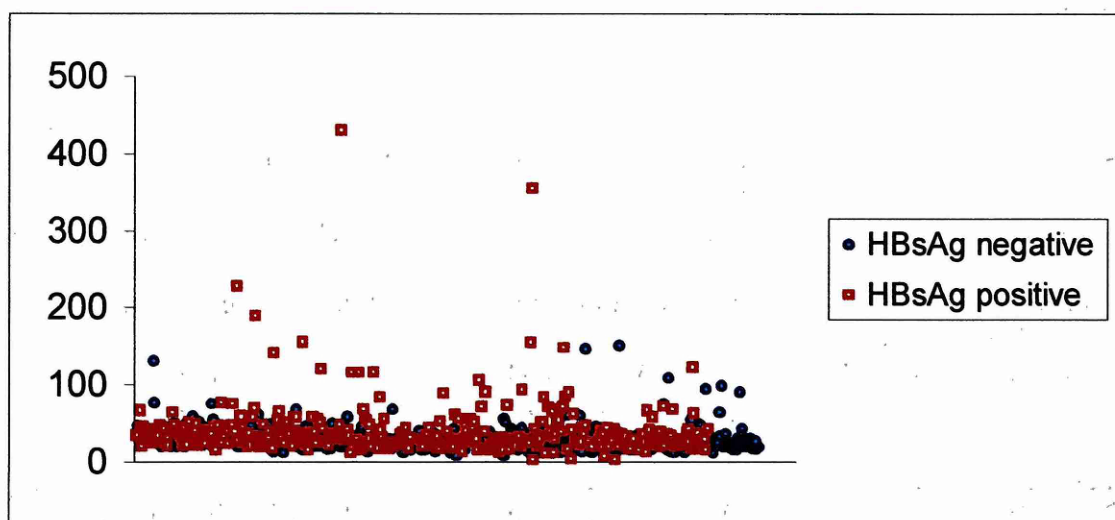


Figure 5. AST levels in HBsAg positive and HBsAg negative patients

AST	HBsAg negative	HBsAg positive
0-39	382	291
40-59	48 (11%)	83 (20%)
60+	16 (4%)	41 (10%)
Total	446	415
Mean age (Years)	30	21

Table 23. Levels of AST in HBsAg positive and HBsAg negative persons

40 IU/ml is widely accepted as cut off value for the test, but Mean +2SD can also be used (ter Borg, 1998). Out of 300 HBsAg negative Gambian samples it was calculated to be 60IU/ml.

AST elevations are significantly higher in HBsAg positive patients ($p<0,001, \chi^2$ test). Interestingly, mildly elevated ASTs are quite common in HBsAg negative persons. Possibly other diseases (malaria) or environmental factors (aflatoxin) could explain this observation. The prevalence of elevated AST in HBsAg carriers is quite high.

AST IU/ml	HBsAg positive		HBeAg positive		HBeAg negative		HBsAg negative	
	Males	Females	Males	Females	Males	Females	Males	Females
0-39	139	151	45	19	89	126	159	223
40-59	50 (23%)	25 (13%)	21 (26%)	11 (30%)	26 (20%)	12 (8%)	24 (13%)	28 (11%)
60+	25 (12%)	16 (8%)	13 (16%)	7 (20%)	12 (9%)	9 (6%)	5 (2%)	7 (3%)
Total	214	198	79	37	127	147	188	258

Table 24. AST levels and HBsAg and HBeAg status in relation to the gender of the patient. Prevalence of elevated AST was similar in HBeAg positive patients of both genders

Clearance of the HBsAg carriage

10% of chronic HBV carriers bled after interval of 6 to 10 years were found to have cleared the HBsAg (see Table 25). All 351 previously known carriers bled during this study were previously confirmed as chronic carriers by double bleeding after a year's interval. No sex difference in this clearance was observed.

Sex	Cleared HBsAg (%)
Males	10 (17/169)
Females	9,3 (17/182)
Total	9,6 (34/351)

Table 25. Relation of sex and clearance of HBsAg over 6-10 year time period

Age at the time of the bleeding	Cleared HBsAg carriage (%)
10-19	8 (18/238)
20-29	9 (5/57)
30-	18 (10/55)
Total	10 (33/351)

Table 26. Relation of age and clearance of HBsAg over 6-10 year time period.

At the time of previous testing, when they were found positive, patients were 6-10 years younger. Data in this table shows that the clearance is increasing after 20 years of age.

4.3 Discussion

The main aim of this study was the collection of human DNA for further genetic studies on families with multiple HBV carriers. In many families a large proportion of the siblings were not bled, because they were vaccinated or not available at the time of bleeding. Children up to 10 years old were not bled, because presumably they were vaccinated. Mother HBeAg status at the time of the actual infection could not be established since the children were bled at least 10 years after the actual infection. Considering all these factors, this study cannot give a complete insight on HBV transmission in The Gambia, but can shed light on the clinical course of the chronic carriage and the epidemiology of infection in families with multiple carriers.

The prevalence of the HBV carriage in siblings of chronic carriers was found to be significantly higher than in the normal population; 37% in sisters against 10% in the normal population and 57% to 15% in brothers. The risk that the first sibling of a chronic carrier will also be a carrier was 45%. The prevalence of the HBsAg carriage in mothers of the carriers was higher than in the normal population (25% to 10%), but the prevalence in fathers was the same, indicating that a mother was possibly responsible for occasional transmission. These data support previous observations of significant clustering of chronic HBV infection in Gambian villages (Whittle et al, 1983 and 1990; Vall Mayens et al, 1990).

The prevalence of anti-HBc, which is a recognised marker of present or past infection, in the Gambian population is higher than 90%, which shows very high risk of exposure and infection (Whittle et al, 1983; Vall Mayens et al, 1990). In our study, the prevalence of

anti-HBc was found to be 93%, suggesting that nearly every person was infected at some stage and therefore was at the risk of becoming a chronic carrier.

Findings of my study suggest that the overall prevalence (10-15%) of the HBV carriage reported in previous studies does not exactly depict the actual situation in the population. Most of the carriers seem to be concentrated in the families. Members of these multiple affected families are obviously at higher risk of developing serious complications. Identification of these families at risk would be an important measure in reducing the costs of the possible therapeutic interventions.

The HBeAg carriage is associated with an increased viral replication, higher viral load, and HBeAg positive patients are more likely to be infectious than HBeAg negative. In the majority of patients, if an infection was acquired in childhood, seroconversion of HBeAg to anti-HBe occurred in adolescent years. This was also apparent in this study, where by the age of 20 to 29 only 10% of the carriers remained HBeAg positive. If we assume that HBeAg positive patients are more infectious then it is clear that in a case of horizontal transmission, infection was acquired from HBeAg positive children or adolescent siblings. That was already suggested previously (Whittle et al, 1983).

This study showed the remarkable difference in HBsAg and HBeAg carriage rates between brothers and sisters of carriers that would mean that brothers of the chronic carrier are at much higher risk of becoming carriers than sisters. The prevalence of HBeAg positives was also found to be significantly higher in adolescent males than in females. The explanation for this difference could either be different X chromosome encoded genetic predisposition or simply hormonal differences. Another less likely explanation could be

that the boys got infected earlier for some unknown reason and therefore have developed HBsAg and HBeAg carriage more often. Rituals such as circumcision and scarification were always considered, but never confirmed as the risk factors (Martinson et al, 1999). Recently published evidence about the high prevalence of HBV DNA positivity in the urine of the chronic carriers (Knutsson et al, 2000) should be considered seriously. Different urination habits for boys and girls could make the difference in transmission pattern.

Another good marker of viral replication is detection of viral DNA in serum by PCR. Although the sensitivity of this assay was improved by optimisation and the data has to be interpreted cautiously, a trend for PCR positivity to decline with years has been observed. Sex difference was very significant at the age group of 15-19, again suggesting that boys clear viremia later than girls.

The high number of vaccinated siblings and siblings not bled did not allow proper analysis of the influence of birth order and family size on the development of carriage. Because of these reasons it is possible to talk only about a trend. In families with carrier mothers the first and the second siblings seemed more likely to become chronic carriers, but that was found not to be statistically significant. In families with multiple affected siblings the prevalence of HBV chronic carriage in younger siblings was significantly higher for the trend when the mother was HBsAg negative or her HBsAg status was unknown, indicating that they could get infected from their older siblings in a very young age. This is consistent with previous reports on birth order as the risk factor for development of chronic carriage and hepatocellular carcinoma (Ryder et al, 1992; Hsieh et al, 1992).

The prevalence of elevated activity of aminotranferases was shown to be significantly higher in HBsAg carriers than in non-carriers. 7% of HBsAg positive patients had elevated ALT and 10% had elevated AST. In HBeAg positives this proportion was higher (13% and 17% respectively). It could be speculated that these persons have more active liver disease but measurements were done only once and the value of single measurement of the aminotransferase levels is uncertain. A rather high proportion of HBsAg negative patients had elevated AST levels (4%). Overall, AST levels in HBsAg negatives were higher than the normal value for the test system (40IU/ml) and a new normal upper value had to be established (60IU/ml =Mean+2SD). Whether the test system was at fault or the Gambian population in general has constantly elevated AST activity remains subject of further exploration. Other factors such as malaria or chronic aflatoxin exposure could possibly cause constantly elevated AST.

The spontaneous clearance of HBsAg has been described previously (McMahon et al, 1990). Our cohort of 351 previously confirmed chronic carriers was followed up over 6-10 years. Overall, 10% percent of the carriers cleared the HBsAg and there was no sex difference in the clearance contrasting with previous reports (McMahon et al, 1990). In the older age group, clearance was found to be higher probably because the majority of patients from this group had already seroconverted to anti-HBe.

5 Molecular epidemiology of HBV in the Gambia.

5.1 Introduction

Over the past decade, increasing attention has been focused on the contribution of variant HBV strains to the clinical course of acute and chronic infection. Mutant HBV could display enhanced virulence with increased levels of HBV replication and resistance to antiviral therapies and vaccine escape.

HBV is a small virus with 4 overlapping open reading frames. The genes are core, surface, X, and polymerase, responsible for encoding core nucleocapsid protein and HBeAg, pre-S1, pre-S2 and S protein, X protein, and protein with priming, RNA- and DNA- dependant DNA polymerase and RNaseH activities respectively.

Seven genotypes A-G of HBV have been identified so far with different geographical distributions (see 2.2.1.2) (Norder *et al*, 1992; Okamoto *et al*, 1988; Stuyver *et al*, 2000).

The envelope proteins are targets of both humoral and cellular response that are involved in viral clearance. Some of the envelope mutants determine the different subtypes of HBV and may also be selected over centuries perhaps under HLA pressure. Changes in amino acids 122 and 160 define the four subtypes: either lysine or arginine at amino acid position 122 or 160 identify d/y or w/r determinants, respectively (see 2.2.1.2)

Little is known about the distribution of HBV genotypes and subtypes in West Africa. Only genotype E has been described so far (Lindh *et al*, 1997), but high prevalence of

genotypes A and D has been reported from North, East and South Africa (Norder et al, 1993).

Surface gene mutants are known to cause breakthrough infections in vaccinated infants and in patients receiving monoclonal antibodies or hyperimmune human immunoglobulin. Vaccine escape mutation from L to G at position 141 has been reported from West Africa (Karthigesu, 1994). A recent study from Taiwan reported increase in prevalence of the *a* determinant mutants in vaccinated population. Gambian mass vaccination was initiated approximately 10 years ago and similar changes in prevalence of mutant viruses could occur.

The core promoter region of the HBV genome regulates transcription of the pre-core and pregenomic mRNAs encoding HBeAg and HBcAg. A double mutation in the basal core promoter (BCP) of the HBV genome appears to arise during chronic infection and is thought to have a selective advantage over wild type viruses due to enhanced viral replication (Kaneko et al, 1995; Kidd-Ljunggren et al, 1995; Laskus et al, 1995).

The presence of HBeAg in the serum of chronic HBV carriers is considered to be an indicator of active viral replication and indicates high levels of infectivity. The frequent virological explanation for the HBeAg-negative profile of hepatitis B are mutations occurring within the pre-core region of HBV DNA which block the translation of the protein. Point mutation, G to A switch at nucleotide 1896 changes the tryptophane (UGG) codon into translation stop codon (UAG) in the corresponding mRNA, is usually responsible for defective HBeAg secretion in these cases (Carman et al, 1989). This

mutation was described previously in South African patients (Kramvis et al, 1997), but has never been previously reported from West Africa.

As an alternative mechanism to abolish the secretion of a viral antigen, target of the immune response, the virus can use mutations within epitopes recognised by T cell response. The frequency of core gene mutations significantly correlates with pre-core stop codon mutation and active liver disease (Bozkaya et al, 1997; Ehata et al, 1992; Carman et al, 1992; Fujiwara, 1998).

In cross-sectional studies mutations were found significantly more frequently in known HBc T helper cell and B cell epitopes (Carman et al, 1995) than in interepitope regions, implying that they could be the result of immune selection or had some growth advantage. Few mutations are noted in the C-terminal portion of the core gene because this overlaps with the highly conserved polymerase gene. New dominant variants with changes within B and CD4+ cell epitopes occurring during the course of disease were detected during occasional exacerbations of the clinical disease in a chronic carrier (Alexopoulou et al, 1997). In the follow up study of anti-HBe seroconverters in remission mutations occurred in the CD4+ epitope, but for patients with ongoing disease mutations were most likely to be found in B cell epitopes (Carman et al, 1997).

Pre-core/core genes from Gambian HBV have never been sequenced previously and prevalence of different variants is completely unknown.

5.2 Optimisation of PCR for the surface and pre-core/core regions of HBV

In order to optimise the sensitivity of the PCR, different sets of primers were initially used.

5.2.1 Surface gene

High enough sensitivity was achieved with the primer combination S3 and S4 (see figure 9), which was shown to work on samples from other parts of the world (P.Karayiannis, personal communication). Hemi-nested PCR was not used, because the task was not to get as many PCR positives as possible, but only to get enough samples for genotyping of the amino acid sequences in The Gambia.

5.2.2 Pre-core/Core genes

With pre-core/core PCR products more problems were experienced.

Set of primers, which had been found to be most effective in studying Greek and Italian patients (P.Karayiannis, personal communication), was chosen.

Primers M3 and Pol 8 were used for the first PCR and primers M3 and

3C (5'- CTAACATTGAGATTCCCGAGA, positions 2439-2460, antisense) for the second round hemi nested PCR.

The sensitivity obtained with M3 and Pol8 (see figure 9) was rather low (20% on HBeAg positive samples) and the combination of M3 and 3C for hemi-nested PCR did not work at all.

A new primer, 5C, was designed (See 3.2.2.3.2) based on sequenced M3-Pol8 fragments from Gambian samples. The combination of M3-5C significantly improved the sensitivity of the hemi-nested PCR (80% on HBeAg positive samples), but it was still unsatisfactory.

The M3-5C (see Figure 9) combination was next used for the first round PCR. The sensitivity improved significantly (80% from the first PCR on HBeAg positive samples). At that stage it could be considered enough for the purposes of this study, but the proportion of PCR positives was very low in the HBeAg negative group, where viral load is known to be much lower and so hemi-nested PCR was employed.

The sequence required for study purposes was already restricted to the minimum with the primer combination M3-5C, so two additional primers 6C and 8C were designed so that both PCR products overlapped each other sufficiently to obtain the full sequence of the pre-core/core region (see Figure 6). Hemi nested PCR with primer set M3-5C (see Figure 10) and 6C-5C achieved 95% sensitivity on HBeAg positive samples and 30% sensitivity on HBeAg negative samples, similar to positivity rates obtained with other patient cohorts (P.Karayiannis, personal communication). M3-8C was less sensitive (25%) on HBeAg negative samples. Samples positive only with both primer sets were later used for sequencing.

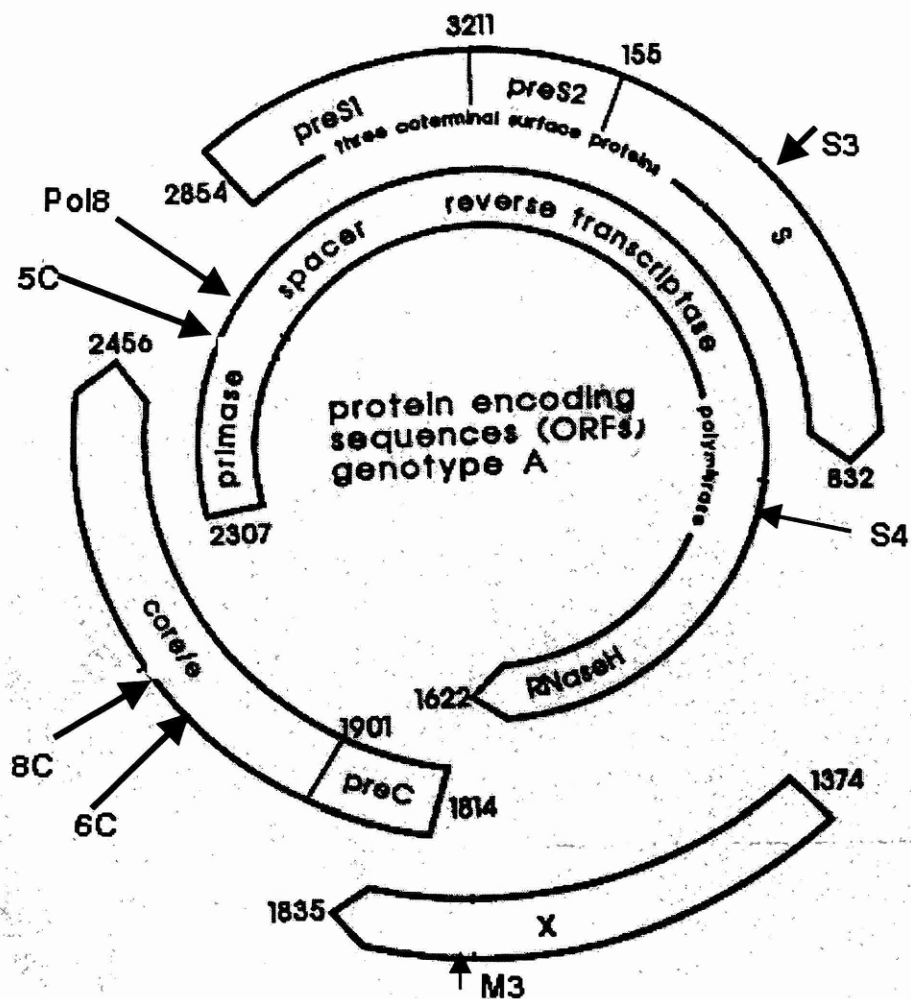


Figure 6. ORFs of HBV and positioning of primers used in this study. (Background picture from Nassal, 1999).

5.3 Validation of PCR results.

DNA extraction and PCR were repeated on 8 samples and the products from both reactions were sequenced to determine the variation of the sequences from two different reactions and establish the probability of mistakes and the risk of possible nucleotide misincorporation. Samples with surprisingly similar sequences from the first sequencing set were chosen. Phylogenetic analysis was also performed to see whether the positioning of the samples in the tree remained the same.

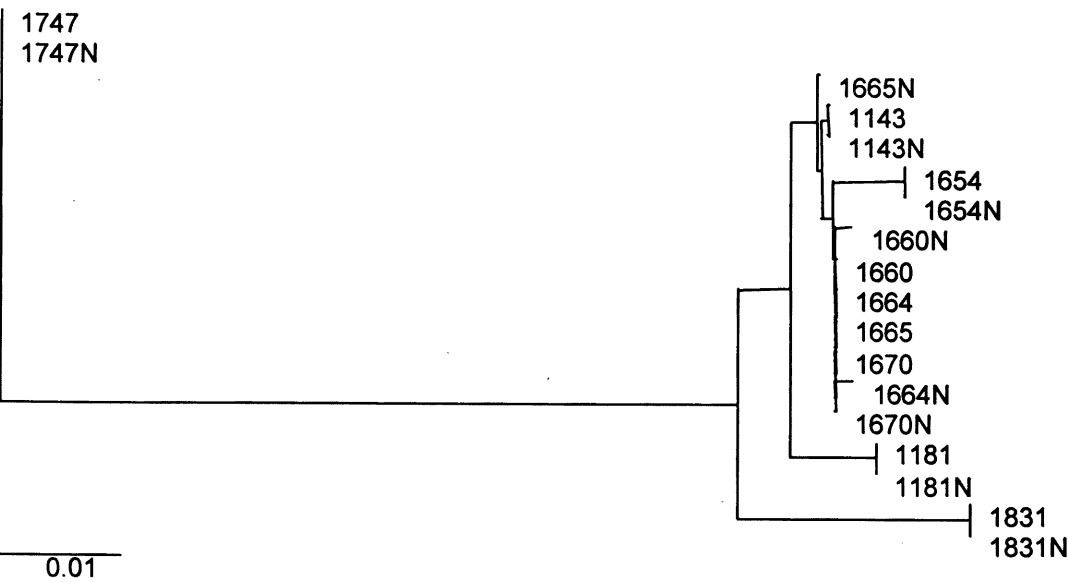


Figure 7. Neighbour- joining phylogenetic tree on pair sequence sets from the same samples. Extension N indicates that sample is from the second PCR set.

On sequencing of the second sample 1665 a small deletion was detected, which could explain the different positioning of (1665N) in the phylogenetic tree. Second PCR products from samples 1664 and 1660 were found to have one nucleotide mutation difference from the first PCR product but did not change their positioning in the tree. All samples were HBeAg positive. This analysis showed that direct sequencing of PCR products was reliable and did not significantly change the positioning of samples in phylogenetic trees (see Figure 7).

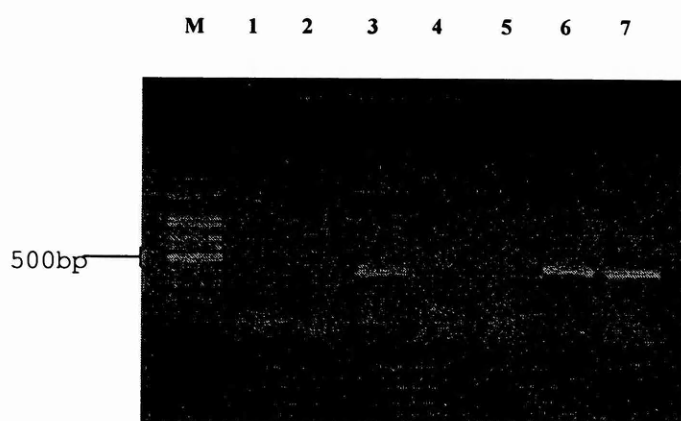


Figure 8. 1% agarose gel with PCR products obtained with primers S3-S4 and stained with ethidium bromide. Legend: M-1,5 kb molecular weight marker, 1- negative control, 2 to 7 -patient samples.

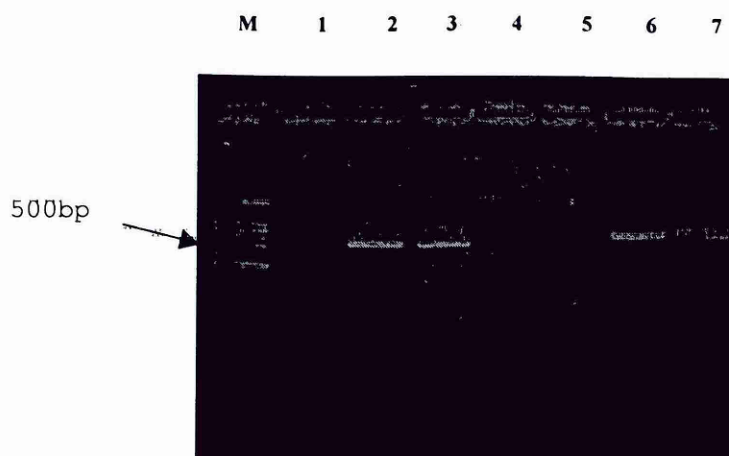


Figure 9. 1% agarose gel with PCR products obtained with primers M3-Pol8 and M3-5C and stained with ethidium bromide. Legend: M-1,5kb molecular weight marker, 1- negative control for primers M3-Pol8, 2 and 3- PCR products with primers M3-Pol8, 5- negative control for primers M3- 5C, 6 and 7 – PCR products with primers M3-5C.

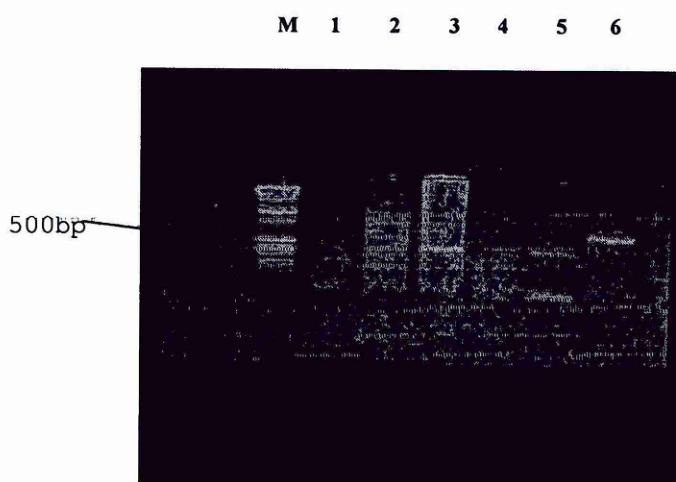


Figure 10. 1% agarose gel with PCR products obtained with primers M3-5C, M3-8C, C5-C6 and stained with ethidium bromide. Legend: M-1kb molecular weight marker, 1 – negative control with the primers M3-8C, 2 and 3- PCR with primers M3-8C, 4 – negative control with primers M3-5C, 5- PCR product with primers M3-5C.

5.4 Sequencing

5.4.1 Genotyping

5.4.1.1 Genotyping by comparing with already published sequences

HBV genotype was determined by comparing amino-acid sequences over the *a* determinant of the S genes of published isolates (Norder et al, 1993). Aligned *a* determinants were compared to published sequences in order to determine genotype and subtype (see Figure 11).

The 6 out of 38(16%) surface gene sequences were found to belong to genotype A. These 38 patients were chosen randomly. Out of 144 precore/core sequences 18 (14%) were found to belong to genotype A. None of these sequences were from the West Region (Basse Area). The rest of the sequences were from genotype E.

Analysis of the genotype of the Gambian samples was first done on sequenced surface gene fragments and then confirmed on pre-core/core gene fragments. The consensus amino acid sequence of the S gene region from genotype A in The Gambia was similar to the published genotype A isolate (*ayw1* subtype) from Cameroon (Norder et al, 1993). The consensus amino acid sequence of the S gene region from genotype E was similar to the published *ayw4* subtype, genotype E sequence from West Africa (country was not stated) (Norder, 1993).

Geno- type	Amino acid sequence
	122 160
E	RTCTTLAQGTSM PFSCCCSKPSDGNCTCIPSSWAFGK
A	K P N T T A 131 140 143 159
B	K P T T A 127
C	K I P T VR 126 160
D	M T T 125
F	K L 158
Gambian A	RTCTTPAQGNSM PFSCCCTKPTDGNCTCIPSSWAFK
Gambian E	RTCTTLAQGTSM PFSCCCSKPS DGNCTCIPSSWAFGK

Figure 11. Comparison of the Gambian consensus sequences from the *α* determinant from genotypes A and E to sequences of other genotypes.

5.4.1.2 Phylogenetic analysis of different genotypes and subtypes.

Phylogenetic tree of the S and Pre-core/Core fragments sequences was generated during the study by including already published sequences from different genotypes and subtypes. The clustering of the sequences into distinct genotypes and subtypes is demonstrated in Figures 12 and 13, constructed with the neighbour-joining (NJ) algorithm.

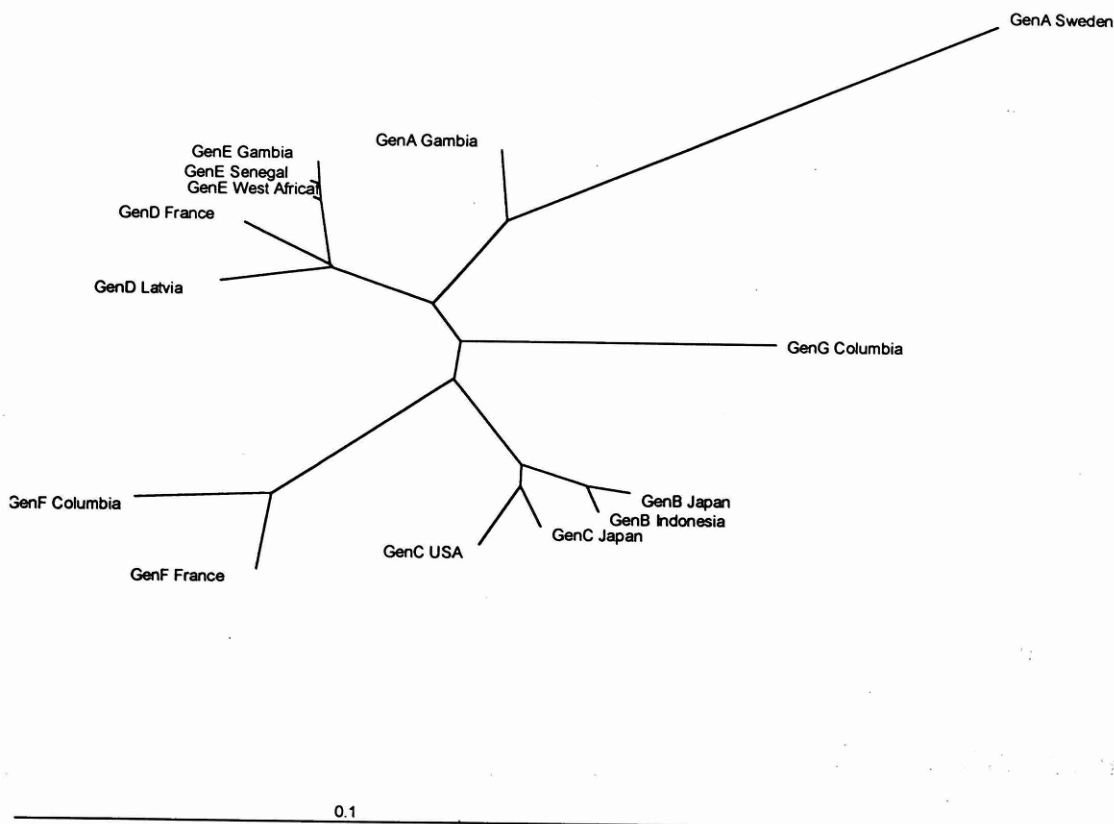


Figure 12. Neighbour joining phylogenetic tree representing sequences from pre-core/core regions from all known human genotypes and two Gambian consensus sequences from each genotype. Designation: Genotype and country of origin.

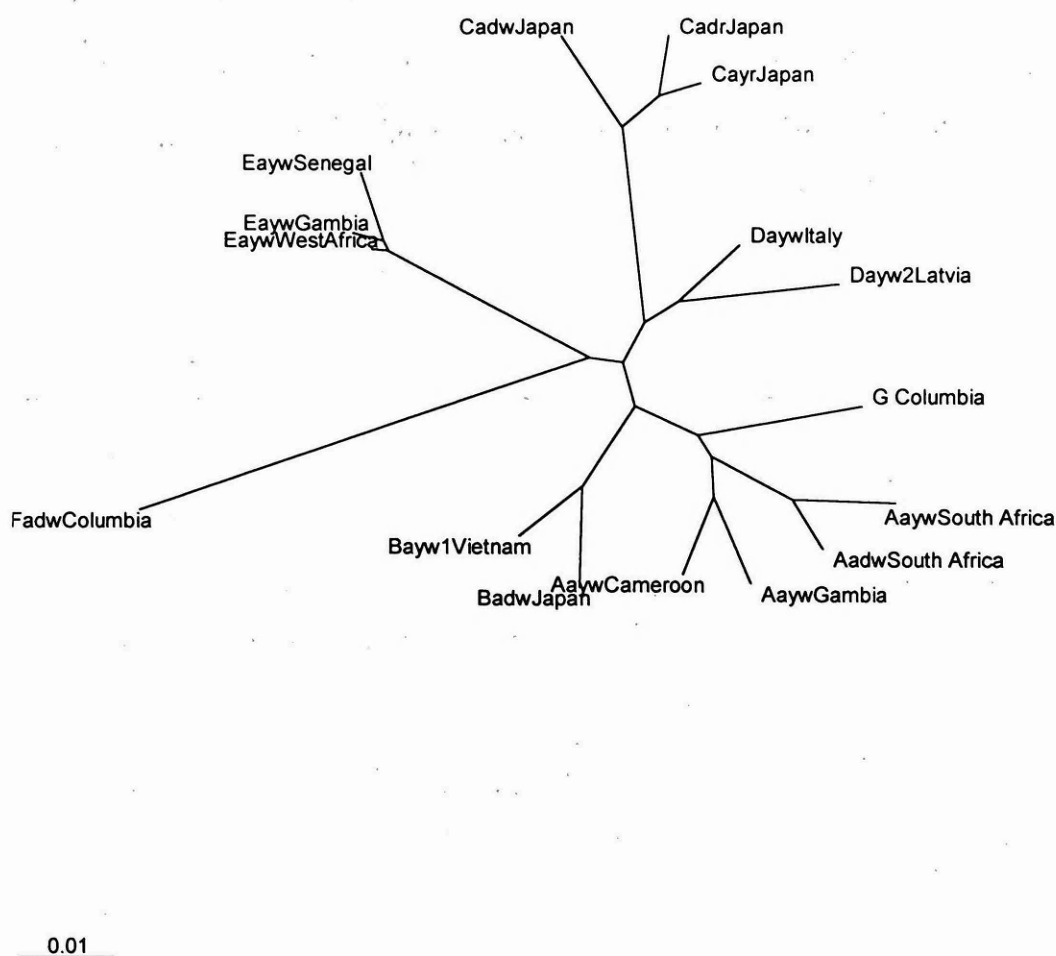


Figure 13. Neighbour joining phylogenetic tree representing sequences from the S gene (nl 509-810) from all subtypes, genotypes and the Gambian consensus sequences from each genotype. Designation: genotype, subtype, and country of origin.

Both figures clearly show, that genotyping can be done either on S or C gene sequences and the result is clear-cut. Separation in subtypes seems to be less significant than division in the genotypes, because the sequences from the S gene clearly cluster according to genotype, not subtype.

5.4.1.3 Geographical distribution of HBV genotypes in The Gambia.

Genotype E seemed to be prevalent around the country, whilst genotype A was detected only in the Western part of the country (see Figure 14).

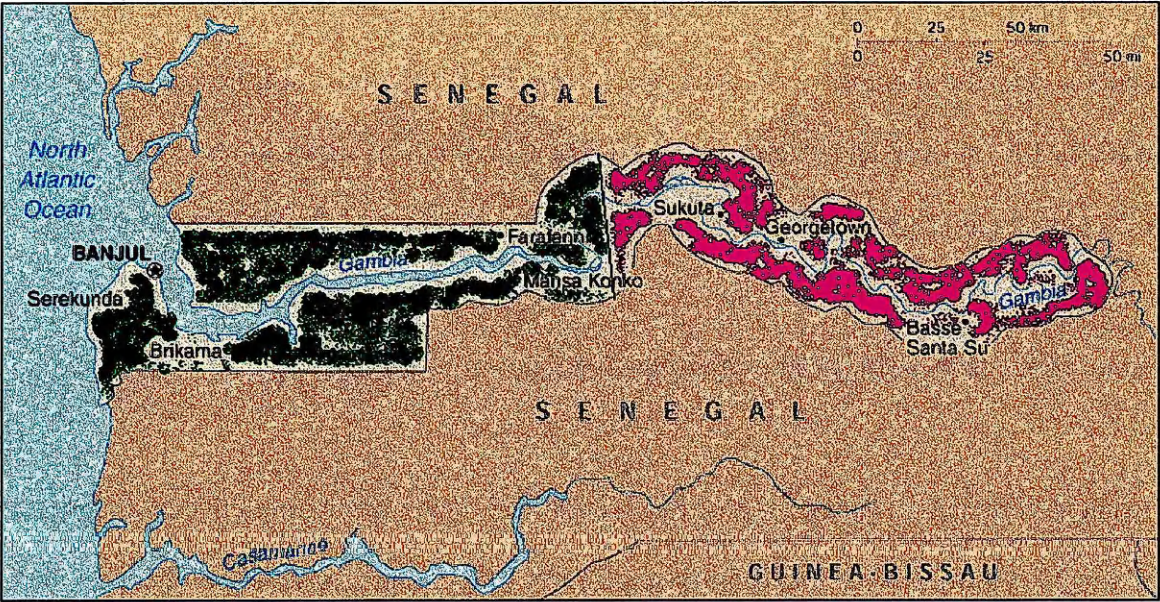


Figure 14. Geographical distribution of genotypes A and E in The Gambia. Designation: green- genotypes A and E, violet- genotype E only.

5.4.2 Mutations in amino acid sequences

5.4.2.1 *Surface gene*

Protein translation from nucleotide sequences encoding amino- acids 112 to 227 of the S gene was carried out in 35 patients. In 8 out of 35 samples single amino-acid substitutions were detected in the surface gene fragment under study (see Figure 15). In four samples amino acid substitutions in the α determinant were observed. In sample 1157, amino acid mutation 123 T to A or T123A was localised in the first loop of the α determinant, has not been reported previously in breakthrough infections in vaccinees or associated with any clinical observations. Mutations in samples 1285, 1144, 1486, 1487 were localised in genotype specific regions. Two genotype E samples, 1285 and 1144, had mutations at positions 140 and 143 and change was from S to T in positions which are normally found in the consensus sequences of genotype A in these positions (see Figure 15). In genotype A samples 1485 and 1486 samples A in position 159 was substituted by G which is a normal amino acid in this position for E genotype.

Premature termination codons at position 216 of the S region preventing full expression of full-length S protein were found in three samples. These mutations had no effect on the amino acid sequence of the open reading frame of the polymerase gene.

5.4.1.5 Core Promoter region

Primer M3 was chosen in order to include the core promoter area. In the 99 samples the sequence of core promoter area was available. In 7 samples (7%) double mutations at nt1762 A to T and 1764 G to A were detected. Four of these 7 samples were HBeAg positive. Two samples were from the mother and her child. This pair also shared the nt1896 pre-core stop codon mutation.

In 31 (31%) samples a mutation from C to T at position 1768 was found. This mutation has never been reported before and did not lead to changes in amino acid sequence. Most likely, it represents the normal sequence of the regional wild type isolate.

5.4.1.6 Pre-core/core genes.

Nucleotide sequence of pre-core region.

All nucleotide mutations found in this study in the pre-core region seem not to affect the secondary structure of the pregenomic HBV-RNA encapsidation signal (see Figure 16)

The association between HBeAg status and the number of mutations in the core region is given in figure 17 and 18. HBeAg positive samples in general did not have any amino acid mutations. Anti-HBe positives on the other hand more often had amino acid substitutions. 12 out of 20 Anti-HBe positive patients were found to carry the 1896 pre-core stop codon mutation. Patients negative for both HBeAg and Anti-HBe who were possibly in a stage of seroconversion window period had an intermediate range of mutations.

All together 14 pre-core stop codon mutants were identified in this study. All these samples were HBeAg negative, 12 samples were Anti-HBe positive and two samples were

not tested for HBeAg markers. In four of these samples an additional mutation at 1899 (G to A) leading to amino acid change from G to D was found.

A mother and her child were found to carry the pre-core stop codon variant. In another family, all three siblings tested were also found to carry the pre-core stop codon variant. Two other samples originated from the same village contained pre-core variants and these may have had a common source of infection.

Gen E

[illegible]

Antie antibody positive

Samples with precore stop codon mutation

All E antigen markers negative

A genotype E antigen positive

Figure 18. All core and pre-core sequences with amino acid substitutions according to their E antigen status. After aa146 core gene amino acid substitutions were not found.

	Genotype E			Genotype A		
HBeAg status	Number sequenced	Number having changes in amino-acid sequence	Avarage number of mutations	Number Sequenced	Number having changes in amino-acid sequence	Avarage no of mutations
HBeAg positive	74	14 (20%)	1.56	14	2 (14%)	1.5
Anti-HBe positive	20	16 (80%)	4	2	0	
HBeAg neg, Anti-HBe neg	20	8 (40%)	2.3	2	0	
HBeAg neg , Anti-HBe not tested	7	2 (29%)	3			
HBeAg and Anti-HBe not tested	5	0	1			

Table 27. Relation between HBeAg markers and number of mutations in the Core gene.

Most of the HBeAg positive samples did not have mutations leading to amino acid changes in the pre-core/core regions. Mutations were distributed over the entire core without an apparent concentration to certain regions, except positions 64 E to D, 93 M to V, and 113 E to D where mutations were recorded more than in one sample. In HBeAg negative samples mutations occurred more frequently. Most of these mutations occurred in B cell epitopes (see Figure 17).

(Carman, 1995)	Chuang, 1993)	(Ehata, 1992)	(Rodriguez-Frias, 1995)	Not previously reported
38 Y to F 46 E to Q 57 Q to H 64 E to D 67 T to S or T to N 69 T to S or T to G 74 V to G 79 P to Q 92 N to H 93 M to V 113 E to D or E to Q 114 T to I 130 P to Q 131 A to P	49 S to C 55 L to I 60 L to E 130 P to Q	49 S to E 55 L to I 59 I to V 60 L to V 74 V to G 113 E to D 130 P to T	12 T to S 27 V to I 116 I to L	5 P to S 13 V to E 14 E to Q 48 C to R or C to H 51 H to Y 68 L to I 100 L to F 125 W to C 125 L to P 146 N to T

Table 28. Amino acid substitutions in the core gene found in this study and in relation to previous reports.

Mutation 64 E to D occurred quite often and it was much more common in HBeAg negative samples with the 1896 stop codon mutation than in HBeAg positive samples

(71% (10/14) and 7% (5/74) respectively). It was also seen in 50% (2/4) of Anti-HBe positive samples without the pre-core stop codon mutation.

A similar distribution of 146 N to T or N to I substitution was found in 50% (7/14) of samples with the pre-core stop codon and in 3% (2/74) of HBeAg positive samples.

Mutations at position 67 from T to S or from T to N I were found only in 50% (6/14) of samples with the pre-core stop codon mutation. Another mutation from 116 I to L or from I to V was found in 29% (4/14) of samples with the pre-core mutation and was absent in the rest of the samples.

Interestingly, in some samples certain mutations diverging from the Gambian consensus sequence are present in other genotypes or subtypes as part of the normal sequence (see Table 30).

Mutations compared to the Gambian consensus sequence	Genotype	Subtype
E genotype (ayw4)		
12 T to S	B and C	adw2, adw4
59 V to I	C	adrq-
64 E to D	D	ayw2, ayw3
66 M to I	D	ayw2,ayw3
67 T to N	B,C	adw2, adr, adrq-
69 T to S	D	ayw2, ayw3
93 M to V	A	adw2
116 I to L	D and E	ayw3, ayw4
116 I to V	D	ayw2
130 P to T	C	adw
130 P to Q	D	ayw2
146 N to T	A, B, C, D, F	adw2, adw, adrq-, ayw3, ayw2
A genotype (ayw1)		
87 N to S	B, C, D, E	adw2, adw, adrq-, ayw3, ayw2
98 I to F	D, E	ayw4

Table 29. Relation of amino acid substitutions found in the Gambian samples to normal sequences in other genotypes or subtypes.

Discussion

A large number of core promoter, pre-core/core and surface gene sequences have been obtained from Gambian chronic HBV carriers. Only three genotype E sequences from West Africa have been published so far (Norder et al, 1993) and information provided by this study is very useful in assessing the predominant strains of the HBV in Gambian chronic HBV carriers.

Core primers, previously found to work on samples from all around the world (Karayiannis, personal communication), did not work satisfactorily on Gambian samples; this was an indirect indication that the nucleotide sequence of these isolates was different. A new set of primers was designed and at the end of the optimisation the sensitivity of the assay was 95% on HBeAg positive samples and 30% on HBeAg negative samples, which was considered as optimal and good for the study purposes.

Two predominant genotypes of HBV were identified in samples from Gambian patients. Genotype E was predominant and sequences from the surface and core genes showed that it was quite similar to previously published isolate from Senegal. Sequences of genotype A were also detected and represented 14% of the total sequenced and the consensus sequence was very close to previously published genotype A sequence from Cameroon. Genotype E was spread throughout the whole country, but genotype A was found only in a Western part (see Figure 14).

The double mutations at nt 1762 from A to T and at 1764 from G to A in the part of the X ORF that contains the core promoter were never previously described in West African

population. In our survey 7% (7/99) of HBV carriers were found to carry this mutation. 4 out of 6 patients with this mutation had elevated AST and ALT levels, significant in comparison to the prevalence of elevated aminotransferases in non-mutant population, a finding that indicates that they could have had higher viral replication rates, as it was suggested previously (Laskus et al, 1995).

The pre-core stop codon mutation also has not been reported previously in West African population. In this study 14 pre-core mutants were detected out of 141 sequences (10%). Pre-core variants were found only in genotype E. Nucleotide sequence substitutions in the pre-core gene were found in 24 out of 144 (17%) sequences. None of these could affect the secondary structure of the pregenomic HBV-RNA (see Figure 16) and therefore HBeAg production.

The findings of the present study indicate that the HBV genome is extremely stable unless exposed to host immune pressure. Analysis of the nucleotide sequences will be reviewed in detail in Chapter 6, dealing with their phylogenetic analysis.

Most of the samples sequenced did not have any amino acid substitutions. Mutations were distributed over the entire core gene without an apparent concentration to certain regions. Slightly more amino acid mutations occurred in B cell epitope regions (see Figure17). Mutations were more frequent in HBeAg negative samples than in HBeAg positive samples. Amino acid substitutions in core gene at position 64 from E to D, 69 from T to S and 146 from N to T were much more common in anti-HBe positive samples with the pre-core stop codon mutation. Why these particular transformations occur mainly in anti-HBe positive patients is not clear.

Although many families were bled, only few of the siblings were found to have the same amino acid mutation. Family clustering was only significant in patients with the pre-core stop codon mutation. Sample 1162 is the sample from a mother and 1158 was sample from her child. They both shared core promoter and pre-core stop codon mutations, but otherwise the amino-acid sequences were very different. Samples 1824, 1825, and 1827 were obtained from siblings from the same family, and 4640, 4690 from non-related children from the same village. They all shared pre-core stop codon mutation, but the sequences had a lot of other different amino acid substitutions. This could reflect how a common variant was selected in different host genetic backgrounds.

In the surface gene only one amino-acid mutation 123 T to A was found which could potentially affect the structure of the first loop of the α determinant, but it has never been previously associated with any clinical significance. The previously published vaccine escape mutation at position 141 was not observed in chronic carriers studied. This indicates that if this mutation occurs, its prevalence is not high in the normal population. In four samples amino acid substitutions characteristic for other genotypes were observed. That was similar to findings in core gene and its clinical significance is not clear.

Premature termination codons at position 216 of the S region preventing full expression of full-length S protein were found in three samples. These mutations did not appear to change the amino acid sequence in the overlapping open reading frame of the polymerase gene. The mutation was previously described only in immunocompromised (Preikstat et al, 1999). Clinical importance of this mutation in Gambian population remains to be investigated.

6 Phylogenetic tree analysis of HBV sequences in Gambian families.

6.1 Introduction.

Most HBV infections in sub-Saharan African infants and children are known to be acquired through horizontal transmission of the virus in early childhood (Whittle et al, 1983; Tabor et al, 1985; Davis et al, 1989). However, the exact mechanisms of spread have never been documented (See Introduction). Significant clustering of HBsAg positive siblings was previously described in Gambian families and transmission of HBV in between them was suggested (Whittle et al, 1983 and 1990).

The organization of Gambian households is quite specific. An extended family is a predominant social structure. In the villages people tend to live in compounds. Usually families with their close relatives share a compound. The number of families in the compound can vary from 1 to 5 or even more in rare occasions. Polygamy is widespread, especially in rural areas. Polygamic families were also included in this study, but constituted only small part of the samples. Non-paternities were excluded by genetic tests during the human genetic study.

To date only a few studies have addressed the question of intrafamilial transmission and clustering of HBV by sequence analysis. In order to determine the extent of intrafamilial versus intracompound or intravillage transmission, I have applied extensive phylogenetic tree analysis to sequences of the pre-core/core and surface genes, obtained following amplification of HBV-DNA extracted from the sera of HBsAg carriers from Gambian families.

6.2 Phylogenetic analysis of surface gene fragments

The NJ tree was constructed on 38 surface gene fragments in order to determine whether there was clustering according to genotypes, geographical area, villages or families (see figure 19).

Two larger clusters are obvious and they characterise two different genotypes. Larger cluster represents genotype surface gene fragments and smaller cluster contains sequences from genotype A. The difference between genotypes A and E are well defined and confirms my previous observation, that detection of different genotypes by using phylogenetic methods is easy and effective.

The sequences of this surface gene region were highly conserved and the difference between sequences was generally very small. 12 sequences were found to be completely identical, what makes more detailed phylogenetic analysis on certain clustering of sequences not feasible.

However, sequences from two siblings in the family 125 were positioned separately from other genotype A sequences. This shows that even surface gene fragments can be family or village specific. Both samples were known to be Anti-HBe positive and to have unique amino-acid mutations (see Samples 1486, 1487 in Figure 19).

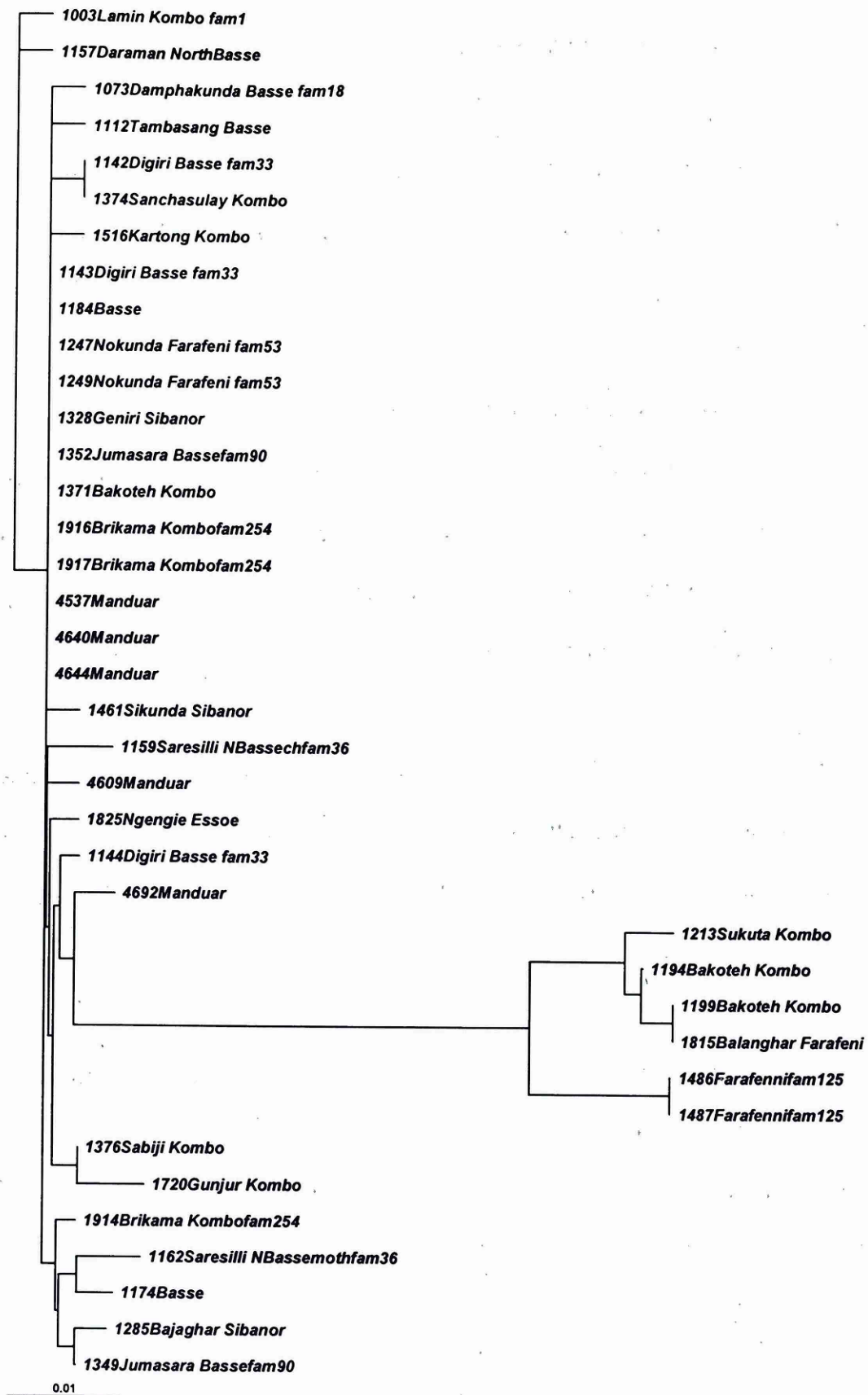


Figure 19. Phylogenetic tree analysis of 38 surface gene nucleotide sequences using the NJ method. Tree is rooted on sample 1003 because of convenience reasons. Taxa descriptions include sample number, village of birth, geographical area and family number if sample was from a family with multiple affected siblings.

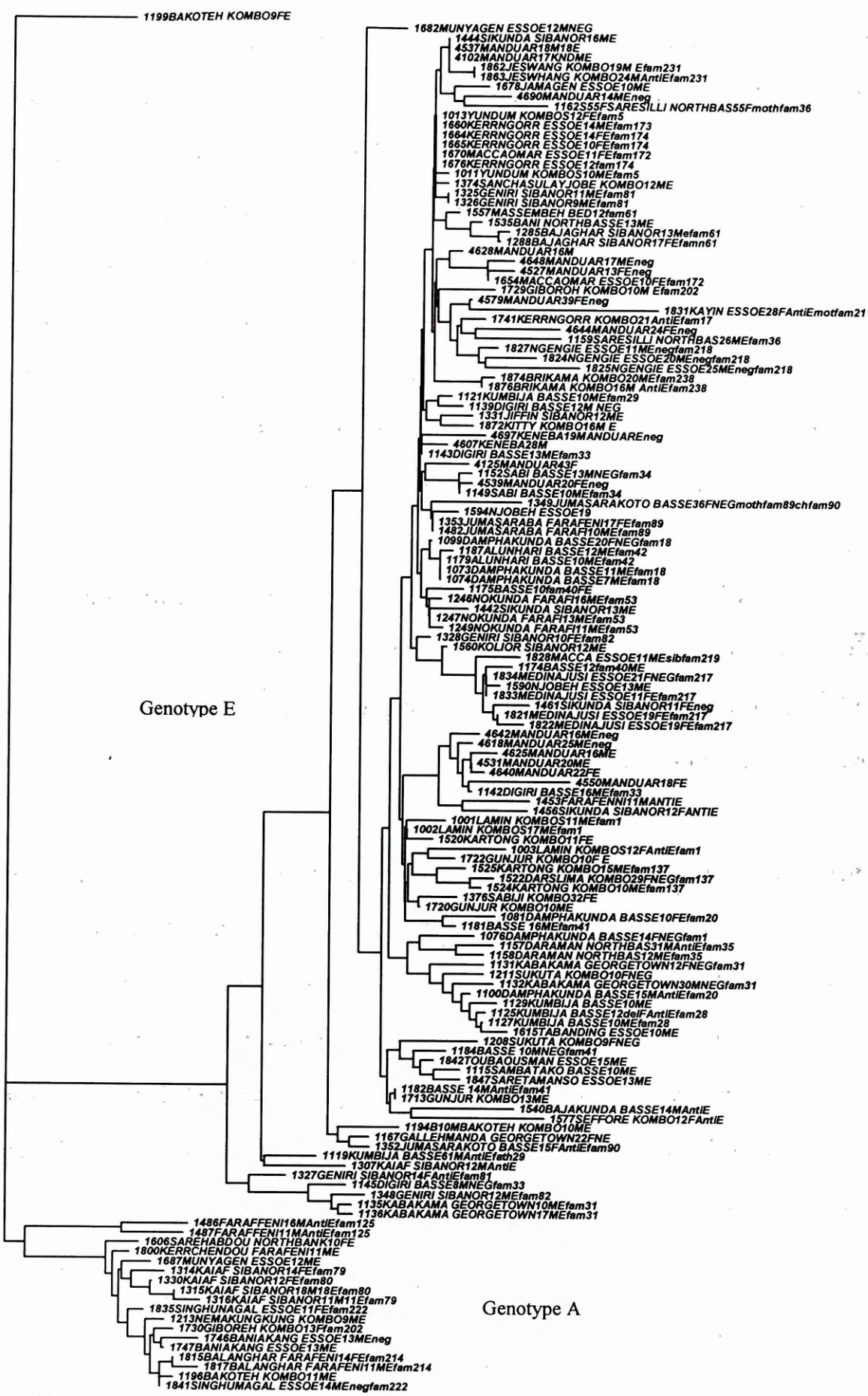
6.3 *Phylogenetic analysis on pre-core/core gene fragments.*

The variability of the pre-core/core gene was found to be higher and therefore phylogenetic analysis for family clusters on the sequence of this gene seemed to be more feasible.

First, a phylogenetic tree was constructed using the simple NJ approach just to see the general trend and positioning of sequences in the tree (see Figure 20). In a case when differences between sequences are expected to be small, the NJ tree is considered to be quite informative and rather similar to the trees obtained with more complicated and perfect algorithms (Maximum likelihood, bootstrap).

The tree was rooted on sample 1199, which was found to be phylogenetically quite distinct. Information about the village of birth, area of the country, age and HBeAg status has been added to the phylogenetic tree in order to see whether there was clustering according to these variables.

Figure 20. Rooted NJ phylogenetic tree on 143 pre-core/core gene nucleotide sequences. Taxa information includes identification number, village of birth, area, age, sex (M or F), HBeAg status (E-HBe antigen positive, AntiE- Anti-HBe positive, Eneg-both E antigen markers negative), and family number, if siblings are from a family with multiple carriers.



The phylogenetic tree in figure 20 revealed the following information:

Two major clusters emerged in the tree and they represent genotypes A and E. This tree clearly shows, that genotypes can be clearly defined by phylogenetic analysis on the pre-core/core genes as well as surface genes (see Figure 19). Genotype E was predominant and variability and phylogenetic distances between several individuals in this genotype was found to be higher, probably due to the larger amount of samples examined. Other explanations, such as longer presence in a certain population or greater variability of the genotype E sequences due to higher prevalence of precode stop codon mutation, could be considered, but could not be proven with the information available.

The clustering of different sequences according to geographical area was not seen, except for genotype A was found only in eastern parts of the country (see Figure 14). Differences between the precore/core sequences were obviously too small to analyse the geographical circulation of the virus in The Gambia.

Definite family clustering of sequences in the phylogenetic was observed. More than one pre-core/core sequence from members of 32 families with multiple HBV carriers were included in this phylogenetic tree. In 22 families sequences from several family members were clustered in the same cluster, which could indicate a close evolutionary relationship. Also siblings from the same family and quite different ages clustered together (e.g. Fam 217), a strong indication of intrafamilial circulation of the virus. Clustering according to village of birth was also obvious in some cases and needed further evaluation.

6.4 Detailed analysis on family and village clustering.

After obtaining the NJ tree, which provided preliminary information, more detailed analysis was performed to confirm family or village clustering of the viral sequences. The presence of different genotypes and also the observation that HBeAg negative samples tend to mutate more due to increased immune pressure was taken into consideration.

To maximise the efficiency of data analysis, only samples with more than one sequence from each village of birth or samples representing families with multiple carriers were included and single samples not belonging to a family or being a sole representative from the village of birth were excluded. This resulted in four different sequence data sets; (I) all available genotype A sequences, (II) all available genotype E sequences, (III) HBeAg negative samples from genotype E, (IV) HBeAg positive samples from genotype E. Phylogenetic relationships among the HBV sequences sampled were reconstructed using a maximum likelihood (ML) method. More details on construction of phylogenetic trees see in Methods section.

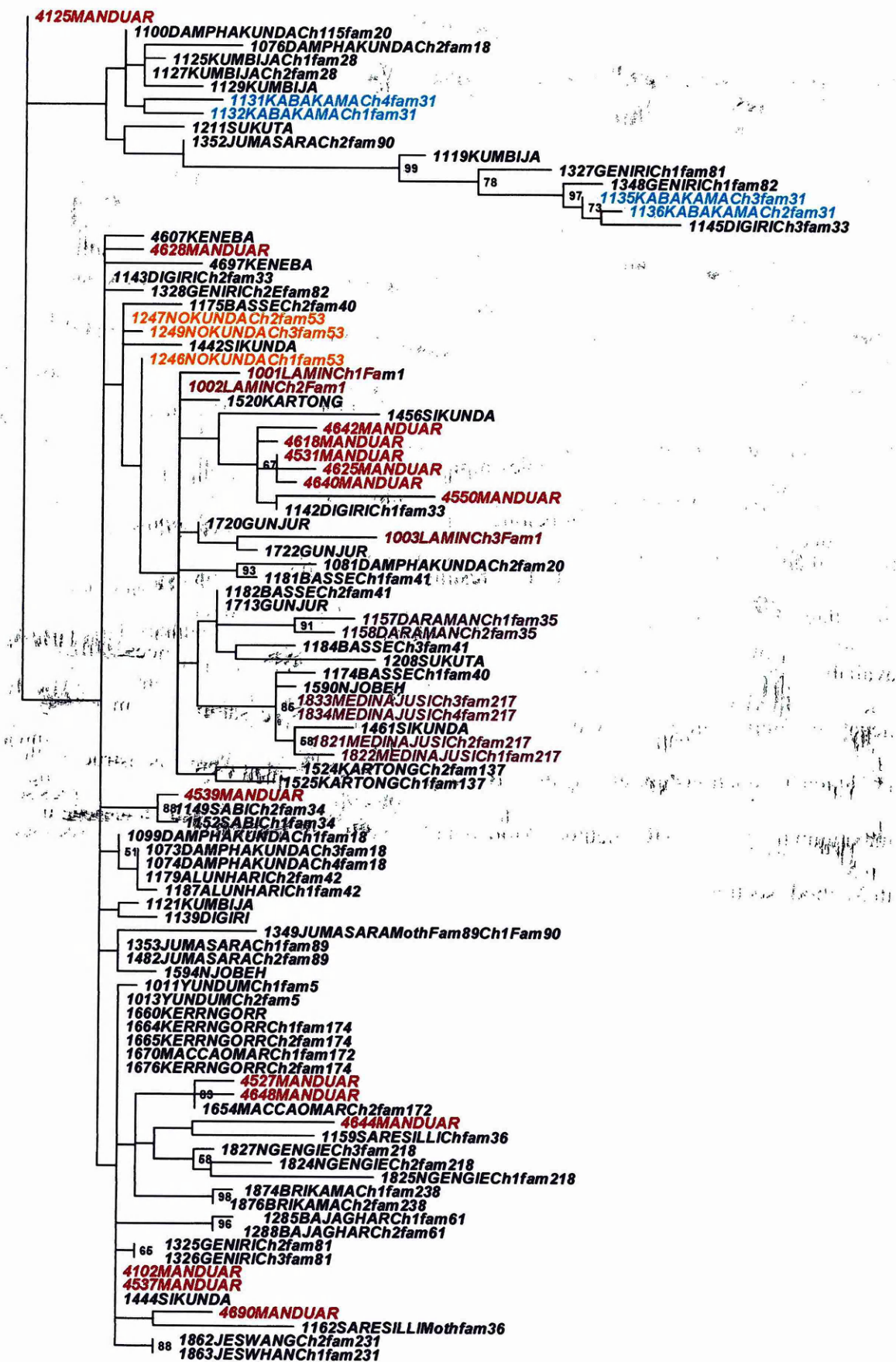
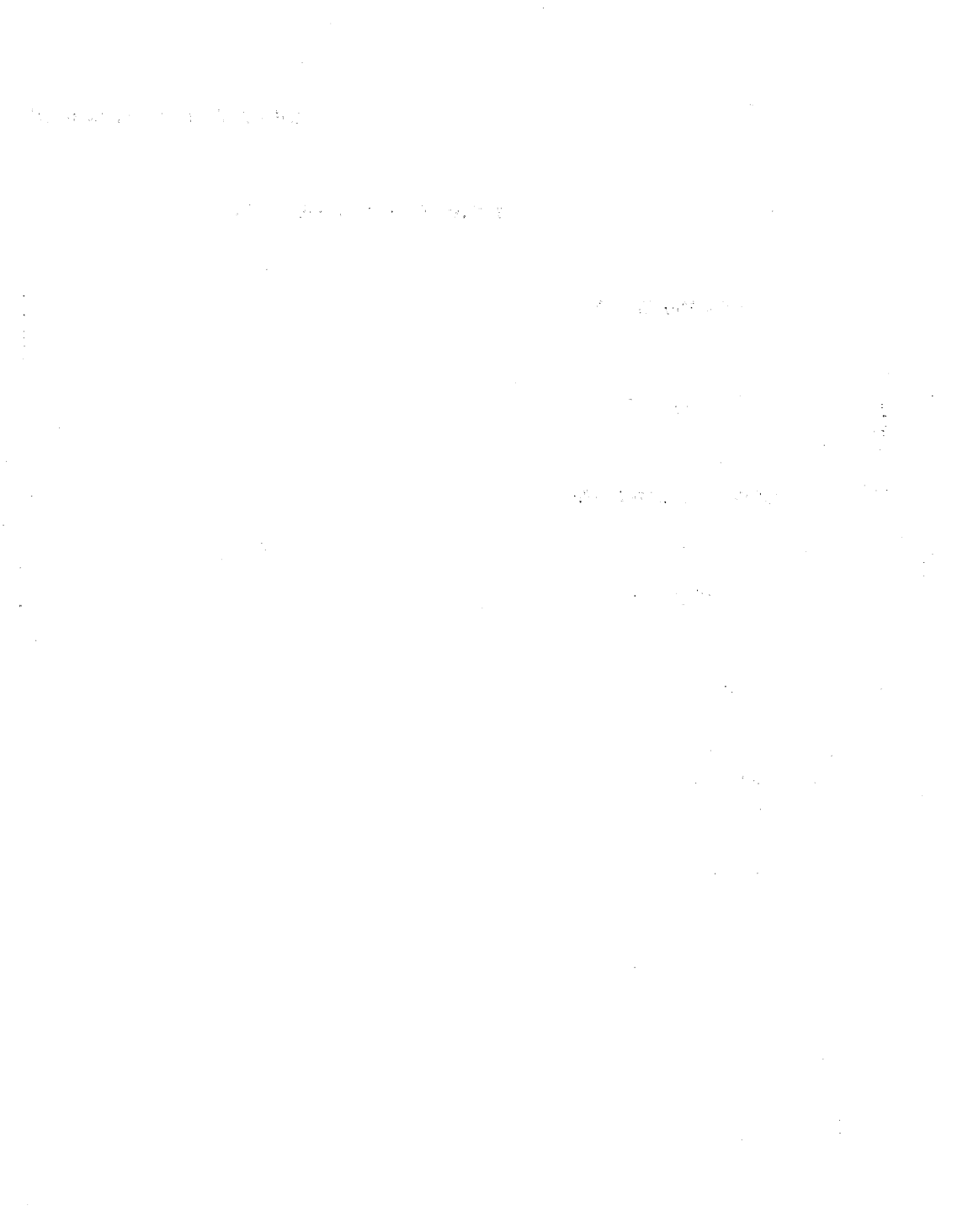


Figure 21. Maximum-likelihood tree ($Ts/Tv=1,60$; $\alpha=0.13$) with NJ bootstrap values for 99 genotype E precore, core gene sequences. Bootstrap values are shown for selected nodes only, where they were higher than 50%. Several examples of village and family clusters are marked with colours.



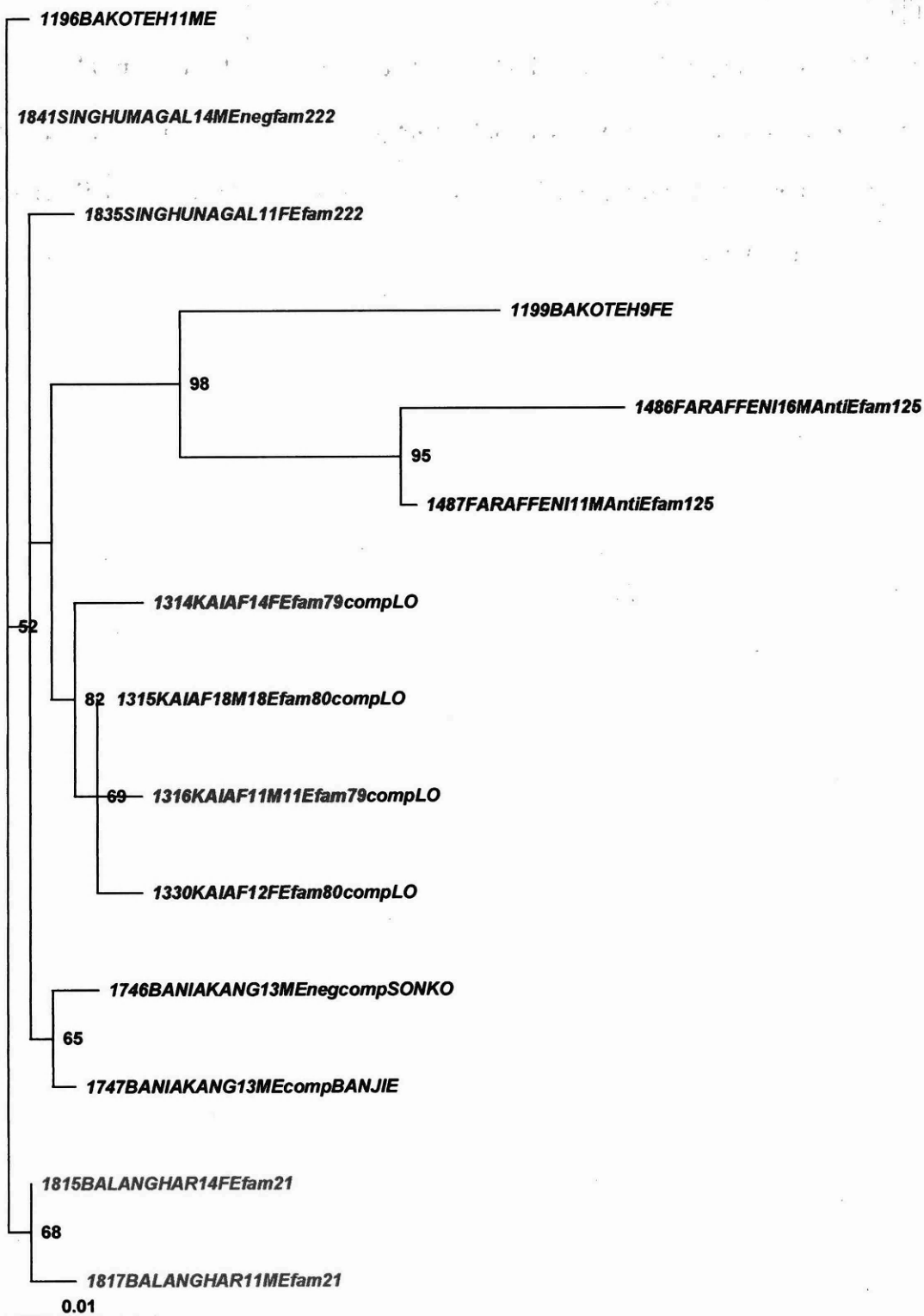


Figure 22. ML tree ($Ts/Tv=1,09$; $\alpha=0.25$) with neighbour-joining bootstrap values for 14 genotype A pre-core/core gene sequences. Bootstrap values are shown for selected nodes only, where they were higher than 50%. Descriptions of taxa include sample number, village of birth, age, sex, E antigen status, family number, and compound name (*comp*) in some samples

The ML trees (with NJ bootstrap values) for the HBV genotype A and E sequences are presented in figures 21 and 22. The general pattern of clustering according to family or village is apparent, although only few nodes are supported by the high levels of bootstrap values as might be expected given the conserved nature of the sequences.

In the genotype A (see Figure 22) most of the nodes are supported by the sufficiently high bootstrap values. Bootstrap values higher than 50% are indicating that node could be of importance (Holmes E, personal communication). Nodes supported by 70% and higher values can be considered as statistically significant.

Families 125 and 214 are located separately with 95% and 68% bootstrap support respectively. Sequences from family 222 cluster together but this cluster does not have sufficient bootstrap support. Of interest is a village cluster from Kaiaf, which has 82% bootstrap support. It is difficult to separate siblings from family 79 and family 80 suggesting that the source of infection from these families was the same or they got infected from each other. Both families inhabited the same compound. Two sequences from the village Baniakang show strong relation with 65% bootstrap support and phylogenetic distances between these two sequences are not longer than distances within

other families. These two sequences were obtained from persons living in different compounds.

In the genotype E the number of samples is much higher and clustering between families and villages is not so well pronounced, but several clusters are still supported with bootstrap values higher than 50% (see Figure 21). Families 231, 81, and 238 are positioned separately in the phylogenetic tree and are supported by high bootstrap values. Some families have very similar sequences to the families from different villages or even different areas, and this most likely is coincidental due to the conserved nature of the sequences in general. For example, families 18 and 42 are positioned in one cluster but are from geographically completely different villages.

In several families and even villages, viral sequences seem to be completely identical. In families 88, 65, 18, 217, 174 sequences from the siblings of the same family are identical. In villages Kerrngorr and Maccaomar members of the families 174, 172 and one single sibling carry exactly the same viral sequence. These two villages are small and geographically very close, so communication between them cannot be excluded. PCR on these samples were repeated in order to exclude contamination artefacts and result was confirmed with the exception of only one nucleotide difference in some samples (see Figure 7). A sequence from family 5 was also identical to sequences mentioned above, but the villages of birth are located very far apart and I cannot imagine any possible relationship between them. Direct communication (travel) seems to be highly unlikely.

Two sequences from the village Manduar were also completely identical. They were obtained from HBeAg positive non-related subjects. Another large cluster of 6 sequences

from Manduar was evident and supported by a 67 % bootstrap value. Children from this cluster are not relatives and do not live even in the same compound. Sequence 1142 also belongs to the same cluster, but represents a completely different geographical area and possible connection between these subjects again seems very unlikely.

Many other families also cluster in family clusters, but these clusters do not have sufficient bootstrap support.

Interestingly, families 40 and 41 do not show any signs of clustering in villages and compounds at all. Samples from both of them were obtained in the regional town Basse, which cannot be considered as a village and has much larger population. Therefore the risk of contracting the infection from somebody outside could be higher.

6.5 *Phylogenetic analysis according to HBeAg status.*

It can be assumed, that most of the patients recruited in this study were infected in early childhood. Over a time period of more than 10 years, a number of new mutations could occur, especially in HBeAg negative subjects. An increased number of mutations and reverse mutations could significantly change the positioning of the sequence in the phylogenetic tree. Separate analysis of HBeAg positives and HBeAg negatives was performed in order to see whether it would make any difference.

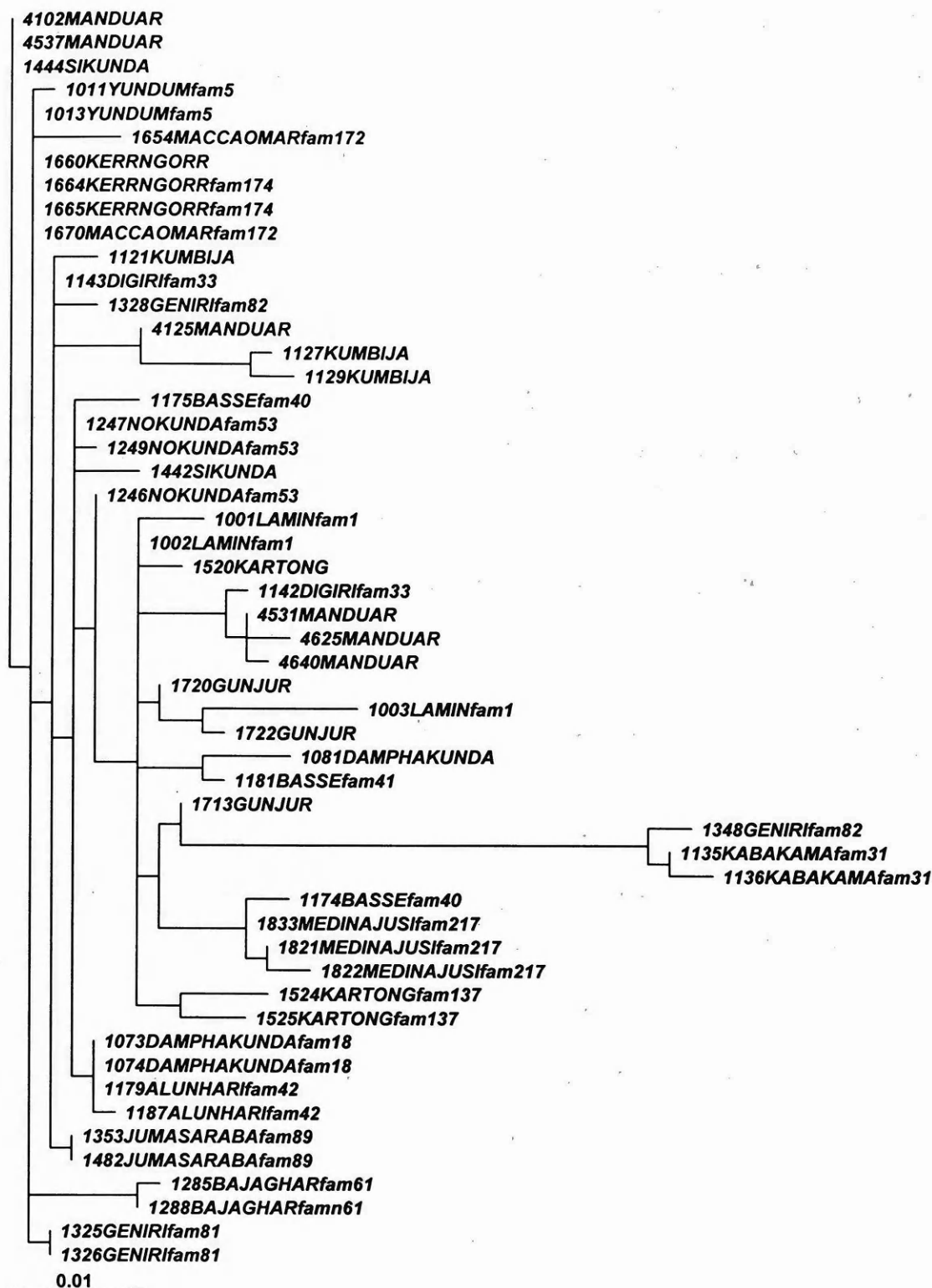


Figure 23. ML tree ($T_s/T_v=1,10$; $\alpha=0.24$) on HBeAg positive samples. Taxon descriptions include sample number, village of birth and family number.

Family and village clusters were more pronounced in the phylogenetic tree in figure 23 of sequences from HBeAg positive subjects than in the common tree in figure 20.

In HBeAg positive samples village and family clusters were definitely more distinct, but some sequences, for example from families 33, 40 and 82 were still scattered throughout the phylogenetic tree and did not show any signs of relatedness.

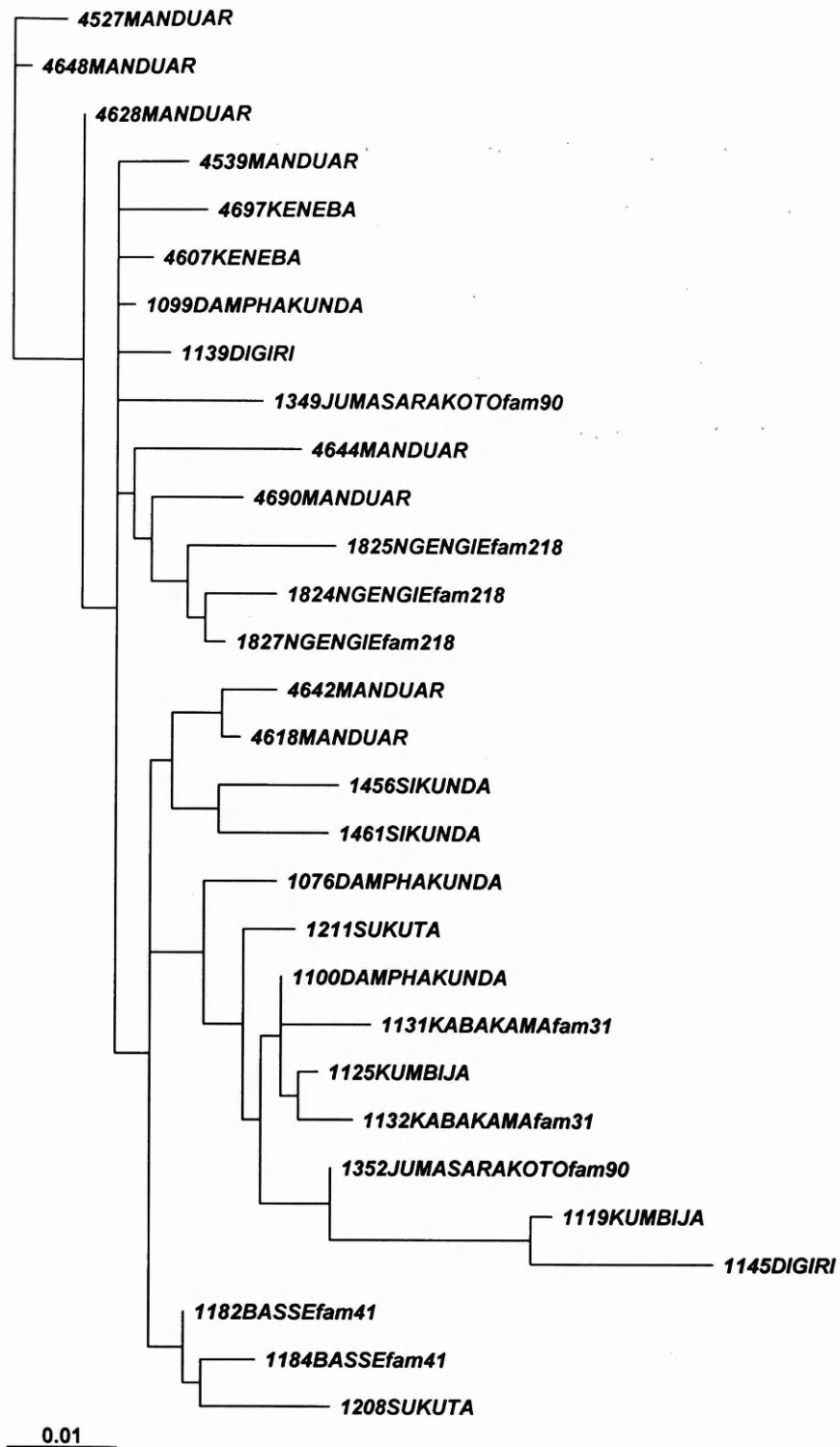


Figure 24. ML tree ($Ts/Tv=1,06$; $\alpha=0.19$) on HBeAg negative samples.

Taxon descriptions include sample number, village of birth and family number.

Family and village clusters were not so well pronounced in the phylogenetic tree in figure 24 of sequences from the HBeAg negative subjects. The branch length overall was longer than that of HBeAg positive sequences, possibly because the increased number of viral mutations in HBeAg negative samples. Interestingly, family 218, which carried pre-core stop codon variant still managed to cluster together in the phylogenetic tree despite large number of different amino acid mutations in the siblings (see Figure 18)

Tree	Number of taxa	Ts/Tv	α
All A genotype	14	1.60	0.13
All E genotype	99	1.09	0.25
HBeAg positive	53	1.10	0.24
HBeAg negative	30	1.06	0.19

Table 30. Summary of data and substitution parameters of the ML phylogenetic trees constructed.

Ts/Tv – *Transition/transversion ratio*

α - *Shape parameter of a gamma distribution of rate heterogeneity among sites*

6.6 Sequences from mothers and their children

Sequences from four mothers and their children were obtained. These were the only PCR positive mothers from those tested. Separate phylogenetic analysis was performed in order to see whether mother to child transmission could possibly occur in this limited number of families.

In the phylogenetic tree shown in figure 25 of positive mothers and their children, only one family (219) did not seem to share similar sequences. In families 137 and 89, phylogenetic clustering was also demonstrated in the main phylogenetic tree (see Figure 20). A mother and her child in family 36 shared both the core promoter mutations and precore stop-codon mutations. Both mother and sibling were already quite old, and viruses probably had many mutations. This could explain the large phylogenetic distance of these viruses and why they did not cluster together in the larger tree (see Figures 20 and 21).

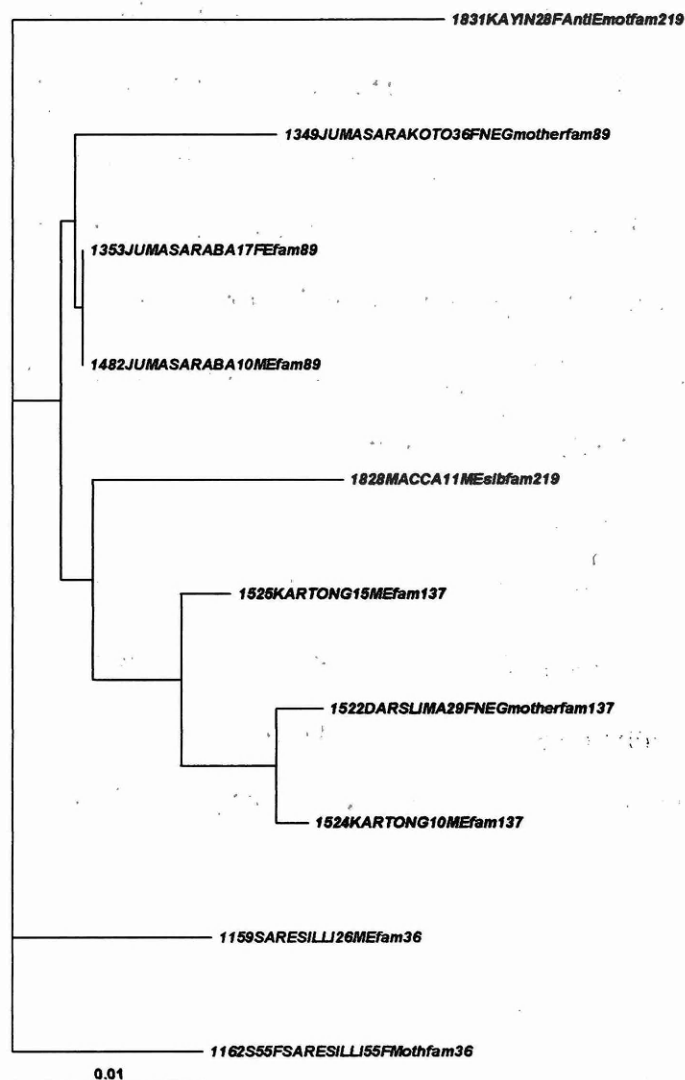


Figure 25. ML tree of sequences from families with a HBsAg positive mother and their children. Families are marked with different colours.

6.7 Sequences from the village Manduar.

16 sequences altogether were obtained from the village Manduar. These patients were not genetically related, although most of them were from families with multiple carriers. Persons bled in Manduar very often were older than 20 years and their viral load was apparently very low, therefore it was very difficult to get a larger number of PCR positive samples for sequencing. Though sequences did not represent relatives and not even the same compounds, most of the sequences were part of several distinct clusters in the phylogenetic trees, mimicking family clusters

Several clusters are well pronounced in the phylogenetic tree in figure 26. Most of them have bootstrap support higher than 50% and even 75%, what indicates that clustering is really significant and viruses within the village are evolutionary distinct. See more comments on clustering in village Manduar under analysis of Figure 20.

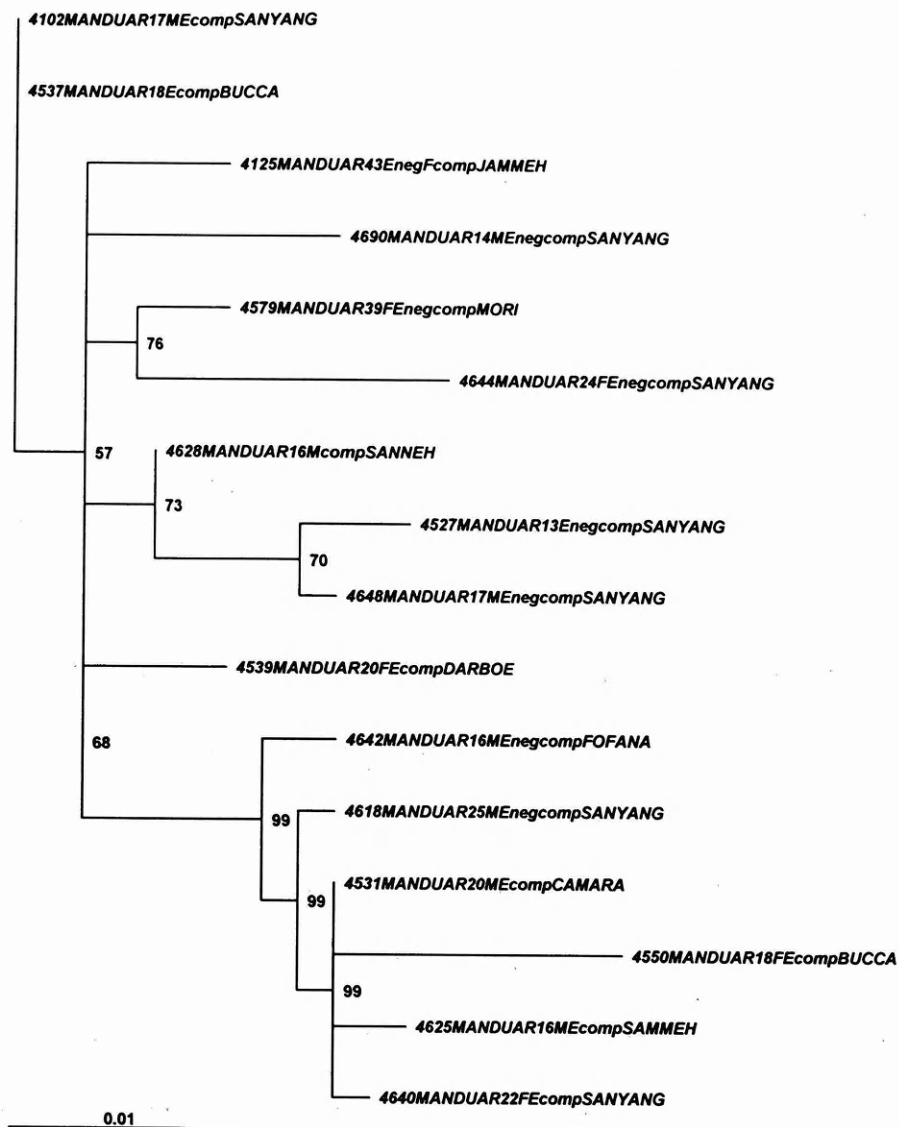


Figure 26. ML tree with bootstrap node values on pre-core, core sequences from the village Manduar. Taxon descriptions include sample number, village of birth, E antigen status (E and E neg), and compound of birth.

6.8 Statistical analysis of the clustering in families and villages

A ML approach was also used to determine whether sequences were clustered by family and village more than might be expected by chance alone. First, for each of the four data sets, 200 random tree topologies were constructed. The log likelihood of each of these replicate trees was then estimated, again using empirical Ts/Tv and α values. As these trees represent random assortments of the data, they effectively constitute a null distribution of log likelihoods. Next three “model” trees, representing clustering at different levels of population structure, were constructed using the TreeView programme: (I) one in which only sequences from the same family were grouped together, with all other branches left unresolved, (the “family” tree) (II) one in which only sequences from the same village were grouped together, with all other branches left unresolved and ignoring what family they came from (the “village” tree) and (III) one in which sequences from the same family and village were grouped together (the “family+village” tree). The log likelihoods of these model trees were then estimated under the data as before. If these likelihoods fall outside the null distribution then we could conclude that the HBV sequence data shows more clustering by family and village than might be expected by chance. It should be noted that this randomisation test is highly conservative because most branches are left unresolved, particularly when the data is only partitioned by family, and therefore ignores any other phylogenetic signal in the data.

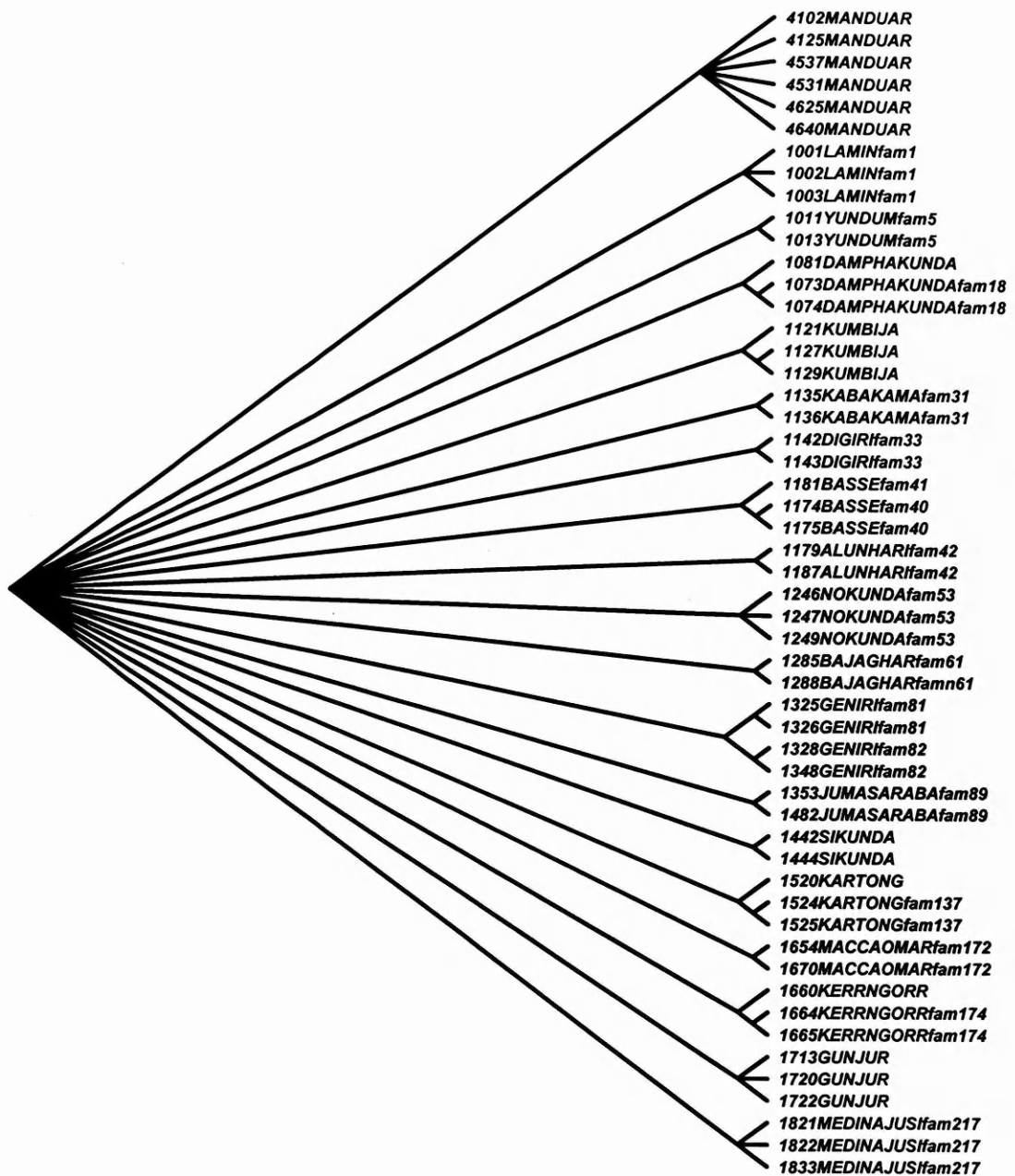


Figure 27. Example of how the maximum likelihood tree from HBeAg positives was transformed to a model tree with combined family+village clustering.

The maximum likelihood randomisation test also provided strong evidence that viruses tend to be related according to the family and/or village they were sampled from (see Table 32). In the case of both the complete A and E genotype data sets the log likelihoods for all three model trees fell outside the null distribution provided by the random topologies, strongly suggesting that there is more clustering by these variable than might be expected by chance. In the case of the large genotype E data set, the combined family+village tree also had a much higher likelihood than either component tree suggesting that these two levels of population structure contain the most information. All three model trees also fell outside of the null distribution of log likelihoods in the analysis of the HBeAg positive genotype E sequences, but not in the HBeAg negative sequences. The latter may be caused by common immunological selection pressures, which increase the extent of convergent evolution.

Genotype	<i>n</i>	Null Distribution	Family Tree	Village Tree	Village + Family Tree
A	14	-1328.12351 to -1389.84444	-1323.49993	-1306.50274	-1306.50274
E	99	-4038.09955 to -4197.28073	-3851.96791	-3816.84344	-3731.26169
E, E antigen +	53	-2202.85834 to -2394.42865	-2188.84172	-2155.96565	-2129.27292
E, E antigen -	30	-2131.48652 to -2213.21954	-2162.7006	-2139.66302	-2139.66302

Table 31. Log likelihoods for three different “model” phylogenetic relationships of HBV sequences (family, village and village+family) and a null distribution constructed using random tree topologies. (See Appendix VIII)

6.9 Discussion

The exact mechanism of HBV transmission in The Gambia is unclear. Significant family clustering of chronic HBV infection has been reported before and this study supported this observation. Analysis of HBV sequences in different family members provided an insight into the transmission of HBV within the family and yielded new information for further epidemiological or anthropological studies.

A large data set of surface, pre-core and core gene sequences was obtained for phylogenetic analysis, representing different families, villages and geographical areas. The use of phylogenetic analysis for confirmation of epidemiological hypotheses is a relatively new approach.

The data collection was designed to run concurrently with collection of human DNA samples for genetic studies. Some of the siblings from multiply affected families were missing at the time of the bleeding. Many were not PCR positive and it was not possible to obtain viral sequences. Another important factor was that the vaccination program was introduced 11 years ago and many siblings from affected families were already vaccinated and therefore most likely HBsAg negative and not tested. In younger siblings we would normally expect higher viral loads and less evolutionary changes. This could make establishment of the source of infection easier.

After a time period of more than 10 years viral sequences were expected to change significantly, especially in HBeAg negative samples. These changes could possibly alter the construction of the right phylogenetic tree and give some samples different location.

Initially, I was interested in three levels of possible clustering: geographical area, village and family, because it was not known how much evolutionary change the viruses had undergone, because even approximate information on the variability of the sequences was not available.

Sequences of the surface gene fragment did not seem to provide the necessary information, because 12 sequences from different families, villages and geographical areas were completely identical. The S gene fragments from family 125 seemed to cluster separately. Probably these samples could represent some different local variant. A clear differences between genotypes A and E were seen (see Figure 19) but otherwise the phylogenetic tree did not seem to provide the necessary information and sequencing of the rest of the S gene products was abandoned.

Phylogenetic tree obtained on the pre-core/core sequences was more informative, because of a larger variability of nucleotide sequences.

Differences between the NJ tree and the ML tree were minimal. This finding supported assumption that the NJ method could be successfully used for creating reliable phylogenetic trees and saving a lot of calculation time. For more detailed analysis, the maximum likelihood tree and bootstrap analysis were considered necessary. Genotypes A and E were very easily distinguishable in the phylogenetic trees (see Figure 20). Thus phylogenetic analysis not only of surface gene sequences but also of precore/core ones can be used successfully for genotype detection.

Overall, pre-core/core gene sequences in The Gambia were very similar. At the geographical area level, it was impossible to see any significant clustering of the sequences. Sequences from completely different villages and geographical areas were sometimes identical. The phylogenetic distance between the sequences was very small and no clear evolutionary pattern emerged. If one could assume that HBV infection arrived in The Gambia relatively recently and spread from one part of the country to the rest, then it would be possible to see certain consistency in its distribution. The phylogenetic tree however, gives more impression that the infection has been present in this area for quite a long time and it is moving in different directions without significant sequence changes. Genotype A seemed to have been introduced later or had some selection disadvantage, which would explain its lower prevalence and limited spread in the country.

One other explanation for this homogeneity of the sequences could be that the number of possible mutations with similar survival capacity due to the overlapping reading frames is rather low and sequences get saturated very quickly. Reverse mutations may take place which further complicates any evolutionary phylogenetic analysis.

In some families (51, 81, 174, 217, 231) sequences in the siblings remained completely identical over the period since their presumed infection. This is consistent with previous observations by other authors (Hannoun, 2000), where children were known to be infected perinatally from their mothers and immune tolerance was caused by HBeAg. In this study some sequences from non-relatives within villages were also found to be completely identical, what has never been previously described. In the Gambian situation, as it was shown by numerous epidemiological studies (Whittle, 1983 and 1990; Vall Mayens, 1990), actual infection obviously occurred during early childhood. All patients with identical

sequences were HBeAg positive, which is consistent with previous observations, that before seroconversion to anti-HBe sequence mutations are rare. The fact, that children from different families, but from the same village of birth carry the same viral sequence, is very strongly suggestive that they got infected horizontally from a common source or from each other.

During the design of the study, it was expected that viral sequences from different families would develop certain family specific mutations due to different genetic backgrounds and immune pressures. However, the observations on the transmission of HBV in the village did not always support this hypothesis. Another explanation for this apparent similarity after a long period of time could be the fact that population in Gambian villages are relatively inbred and have very similar genetic backgrounds.

In many families and villages the sequences formed clusters which were supported by high bootstrap value ($>50\%$ or even $>70\%$), widely accepted as reliable criteria for the evaluation of the phylogenetic trees. Interestingly, some of the village and family clusters were extremely similar. For example, in the cluster from the village Kaiaf, sequences from families 79 and 80 were clustered together and these clusters were indistinguishable from each other, giving the impression that members of these families got infected from each other or from other common sources in the village. Both families were unrelated but inhabited the same compound.

16 sequences from non related subjects were obtained from the village of Manduar and it was obvious that they represented different phylogenetic clusters. The majority of

sequences were from children from families with multiple carriers, but unfortunately it was impossible to obtain more than one sequence from one or each family. These clusters are nearly indistinguishable by their size and phylogenetic distances from family clusters in other phylogenetic trees, and many of them have significant bootstrap support (>70%) in separate phylogenetic trees from this particular village (see Figure 26), indicating that individuals got infected from different sources. Clusters were representing children from approximately the same age group, but one has to admit, that nearly all persons bled were from the same age group. The village is rather small (500 inhabitants) and isolated from the main roads. Most of the people in The Gambia live in households called compounds. Knowing that, the most probable assumption would be that children get infected horizontally from other siblings or compound neighbours, because that is the place where they spend most of the time. Information about compounds of birth of the children in this village was available. Surprisingly, clusters in the village Manduar did not support this assumption. Most of the individuals were from different compounds and still carried very similar viruses.

Other findings from this study are similar to findings in Manduar. In village Baniakang two non-related subjects from different compounds had very similar sequences (see Figure 20). In village Damphakunda three samples formed part of a family cluster, but other sequences from the village were phylogenetically very distant (see Figure 21). Village Kumbija had a family cluster and two other sequences were very distant (see Figure 20).

The similarities of sequences between several mothers and their children could indicate that mother to child transmission occur in The Gambia, though it is not widespread. Perinatal transmission has been excluded by multiple epidemiological studies (see 2.3.1.4)

so apparently horizontal transmission or transmission through mother's milk is more likely to take place. If an infection is acquired from the mother, then it happens in an early childhood and therefore the virus in mothers since the actual time of infection can have more mutations than in their children, especially in HBeAg negatives. This could explain why sequences from mothers 1522 and 1349 are evolutionary more distant than those of their siblings. They probably shared the same viral strain at the time of infection, but then the virus from the mothers underwent more changes, because mothers were of later became HBeAg negative.

From these limited data one could assume that there is a clear division between different viruses within the village not always reflecting division in families or compounds.

The PAUP 4.0 programme package provides the tools for random simulation. 200 hundred random trees were simulated on each of the four data sets and model trees with all sequences organised according to village and family clusters, and their maximum likelihood log values were obtained. If there was no clustering in the tree, then artificial improvement of the tree according to the rules of clustering would just deform the tree and make it even more similar to any of the random trees, so it would be impossible to tell the difference. In contrast, data sets of the log likelihoods for three model trees fell well outside the null distribution provided by the random topologies. The only tree which fitted within the random topologies was the tree constructed from HBeAg negative genotype E samples. In the case of the E genotype set, the combined family+ village tree had much higher likelihood values than either component tree, suggesting that these two levels of population structure contain the most information. This statistical analysis is supporting the idea that actual clustering is somewhere higher than simple family clustering.

In conclusion, it is clear that phylogenetic analysis can be a very effective tool for epidemiological HBV transmission studies and in addition can raise additional questions for further epidemiological exploration.

7 General Discussion

Only a few studies on HBV transmission within the families have been published so far. Nucleotide sequence comparisons have been used previously to demonstrate transmission of HBV from mother to infant, but also in common source outbreaks of infection in different epidemiological settings (Barlet et al, 1994; Acarca et al, 1994; Santantonio et al, 1997; Ho, 1995).

The availability of a large number of specimens of all ages from the Gambian families with multiple HBV chronic carriers has provided unique opportunity to apply human and molecular epidemiological methods for the analysis of transmission of the virus within the families.

The original aim of the sample collection for this study consisted in the accumulation of human DNA probes from multiple affected siblings for further human genetic studies. Overall, 142 families with multiple HBsAg carriers were identified. Although I have tried to get as many carriers as possible from each family, several factors reduced the number of samples available. A large proportion of the siblings were not bled because they were not present at the time of the bleeding or refused to take part in the study. The Gambia is one of a few African countries where mass vaccination has been introduced in 1986, and vaccination coverage was thought to be higher than 90%, so, the children under approximately 9 years of age were most likely protected against chronic infection (Viviani et al, 1999; Inskip et al, 1991) and therefore not bled. The extent of mother to child transmission could not be established, because the HBsAg and HBeAg status of mothers at the time of delivery could not be known. Considering all these factors, the subjects of this

study represent only a part of the families with multiple HBV carriers, making a full human epidemiological analysis incomplete. However, one could try to characterise the course of the chronic carriage and give an insight into epidemiology of HBV infection within the families. Fortunately, several human epidemiological studies were performed in The Gambia previously (Whittle et al, 1983; Whittle et al, 1990) and had provided some of the missing information about the clustering of HBV infection in Gambian families.

Although a large number of samples have provided good material for molecular epidemiology studies, certain insufficiencies in the original design made the data set incomplete for this particular study. The possibility of using phylogenetic methods for the current molecular study was not considered at the time of the data collection. Patients were recruited from all around the country and in many cases only one family represented the whole village. Ideally, for molecular epidemiology studies of this kind, a large number of samples from a single village would be required in order to better answer questions on HBV transmission patterns in the Gambian households.

Many samples were PCR negative and it was not possible to obtain viral sequences. That reduced the number of families having multiple carriers with available nucleotide sequences to 32. In very young children we would normally expect them being HBeAg positives with higher viral loads and less evolutionary changes in viral sequences, what would make the sample size bigger and transmission study easier. In my subject group, because of the mass vaccination programme, all siblings were older than 9 years and some time after infection could undergo significant changes, especially in HBeAg negative individuals, making analysis of transmission more difficult.

Another problem with this study is its retrospective nature. The data collection took place in The Gambia, but this current analysis was done in the UK and it was not possible to go back and test one or other hypotheses or get more precise information on certain family clusters or related epidemiological information. The amount of blood in the samples was sometimes insufficient for doing all the necessary serology tests due to the difficulties of bleeding in the field surroundings. For these reasons, several samples were not tested, for example, on the HBeAg or aminotransferase activity.

Nevertheless, the large sample size, the unique opportunity for a long term follow up of chronic HBV carriers, the limited information on Gambian and West African sequences in general, and the use of novel and experimental phylogenetic methods provided a new and interesting information on previously unknown molecular aspects of HBV epidemiology in the Gambian families and villages. The detailed phylogenetic approach was shown for the first time to be a very effective tool for the analysis of HBV transmission within families even being used long time (10-30 years) after the actual time infection.

Overall, the prevalence of HBsAg chronic carriage in The Gambia was previously reported to be around 10-15% (Whittle et al, 1983; Viviani et al, 1999; Inskip et al, 1991). The prevalence of HBsAg carriage in my study was 37 % in sisters and 57% in brothers of chronic carriers, that is, well above usual. The risk that the first sibling of the chronic carrier bled would become a carrier in this study was calculated to be 45%. This shows clearly, that the infection is more likely to cluster in families, as already suggested in previous studies in The Gambia (Whittle, 1990), where the chance of the youngest child in household being a carrier of HBsAg was strongly related to the number of antigen-positive

siblings. This cannot be attributed to the increased risk of exposure, because the prevalence of HBV infection (anti-HBc positive) in The Gambia is very high (>90%) and nearly everybody gets exposed to the virus in early childhood. Based on findings of this and previous studies, we can assume that most of the carriers in The Gambia are members of carrier families. That would mean that, the overall prevalence of the carriage (10-15%) does not really depict the real situation in the population. Those at risk of developing serious complications as cirrhosis and HCC are most likely members of these families. Identifying these families at risk would be an important measure for better targeting of therapeutic interventions.

Remarkable differences in carriage rates were found between brothers and sisters of the chronic HBV carriers and between males and females in general. That could mean that brothers of the carrier are in higher risk of becoming carriers. This difference could be explained by different X chromosome encoded genetic predisposition or hormonal differences. Also, if boys would get infected in earlier age than girls, that could lead to higher carriage rates. All boys get circumcised in The Gambia and different urination habits can also be discussed.

Birth order has been suggested previously as an important risk factor for developing chronic HBsAg carriage and hepatocellular carcinoma (Whittle, 1983; Ryder, 1992; Hsieh, 1992). This study confirms previous observations that younger siblings in the families with older sibling carriers are at high risk to become carriers. This trend was found to be significant in families where mothers were known to be HBsAg negatives or their HBsAg status was unknown. They apparently got infected earlier in life from their siblings and therefore a higher proportion of them developed chronic carriage. In families with HBsAg

positive mothers, the higher prevalence of carriage in younger siblings was not observed. It seems that in such cases older siblings got infected from their mothers. Possible mechanisms for that will be discussed later in this chapter.

There is one very important aspect, which was not considered at the time of the study design. Younger vaccinated siblings in families with multiple carriers apparently were exposed to infection from the very early childhood. Therefore, they could be potential carriers of the vaccine escape mutants that could have far reaching epidemiological implications. That is one of the reasons why families with multiple carriers should be closely monitored in the future.

The HBeAg carriage and PCR positivity are associated with increased HBV replication and higher viremia levels. Therefore, HBeAg and PCR positive patients are more likely to be infectious. Prevalence of HBeAg positives in persons older than twenty years in this study was found to be significantly lower than in children and adolescents. Therefore one can safely assume that if a horizontal transmission is the main route of infection in The Gambia than the main sources of infection are children and adolescents. That in all respects supports hypothesis on intrafamilial transmission between siblings.

Prevalence of HBeAg and PCR positives is significantly higher in males than in females and that could indicate that the virus has more active replication in men. This difference was very obvious in the 15-19 age group, suggesting that boys clear viremia later than girls and therefore remain as a potential source of infection longer. This observation can lead to speculation that boys during their lifetime infect more younger siblings or other children

than girls. That would be nearly impossible to prove by conventional methods and a new cohort study had to be designed.

10% of the carriers followed up for a 6 to 10 year period cleared the HBsAg carriage that is consistent with previous reports (McMahon et al, 1990). In elderly this clearance was even more common. I cannot say whether these individuals completely cleared the virus from their liver and blood because more detailed investigations on samples were not performed. Most likely viral replication diminished over the time period and HBsAg became undetectable. Possibility that these individuals could develop mutants that could escape the conventional HBsAg detection tests could also be considered and PCR on samples performed.

Only three genotype E sequences from West Africa have been published so far (Norder, 1993) and information provided by this study was very useful in assessing the predominant strains of the HBV and the molecular epidemiology of the virus in Gambian chronic carriers. Two genotypes of HBV were identified in The Gambia by this study. Genotype E was found to be predominant in the country. The primary structure of the genotype E HBV C and S genes was quite similar to the sequence of the previously published isolate from Senegal (Norder et al, 1993). Sequences from this genotype were obtained from samples all over The Gambia. Isolates of genotype A represented only 14% of the total number of sequenced samples. The primary structure of genotype A representatives was very close to a genotype A sequence from Cameroon (Norder et al, 1993). However, HBV genomes belonging to genotype A were found only in Eastern parts of the country (see Figure 14). A possible explanation for these differences could be that genotype E is older and endemic in the area, whilst genotype A was brought in later and that is why it is spread predominantly

along the coastal area, where main trading routes are located and migration of people is historically greater.

Sequencing provided very important information about predominant nucleotide and amino acid variations in the HBV genomes occurring in the Gambian population. Core promoter, pre-core, and core sequences of HBV have never been obtained from The Gambia and therefore the prevalence of previously reported important amino acid substitutions was unknown. The double mutations at nt 1762 from A to T and at nt 1764 from G to A in the part of the ORF X that covers the core promoter was previously associated with more active liver disease (Okamoto et al, 1994; Hannoun et al, 1997). 7% of Gambian carriers were found to carry this mutation and 66% (4/6) of them had elevated aminotransferase levels. A mother and her child were found to share this mutation, which could indicate possible transmission or simply common genetic background.

The pre-core stop codon mutation has been thoroughly studied and was known to be quite common in Southern European populations. This mutation has not been reported previously in HBV genomes of genotype E and in West African population. In this study, 14 pre-core mutants were detected out of 141 sequences (10%), which is considerably lower than in some parts of the world (Southern Europe, Asia). The overall prevalence of the HBV pre-core mutants may be even lower, since two mutants were found in the mother and her child, three mutants were found in three siblings from the same family and two mutants were in samples from one village. Sequences from these viruses were phylogenetically very similar and that could indicate transmission of the mutant viruses. Therefore, family clustering of pre-core variants of HBV genomes was obvious. Transmission of pre-core variants, even in childhood, was previously associated with

development of fulminant hepatitis in neonates (Sterneck et al, 1998). In Gambian situation, it was very unlikely, that children could develop severe hepatitis and survive, so most likely they did not develop clinical disease, if they were infected by the HBV variant with pre-core stop codon. Unfortunately, my sample collection was prepared 10-15 years after the actual time of infection and more detailed retrospective investigation is impossible. Pre-core variants of HBV genomes were found only in isolates belonging to genotype E. It was previously suggested that these mutants are observed quite rarely in genotype A and no variant was found in our samples.

Most of the HBV samples sequenced did not contain any amino acid substitutions. In the core gene, mutations were distributed over the entire gene without an apparent concentration in certain regions. Substitutions occurred significantly more frequently in the HBeAg negative than in HBeAg positive samples. The stage-dependent mutation rate in chronic hepatitis B, with few mutations in the HBeAg positive tolerance phase and increased number of mutations in the anti-HBe phase, has been previously demonstrated in several cross-sectional studies (Akarca et al, 1997; Karasawa et al, 1997).

Interestingly, amino acid substitutions in the HBV core gene in position 64 from E to D, 69 from T to S and 146 from N to T were much more common in the anti-HBe positive samples with the pre-core stop codon mutation. All these substitutions have been previously reported as normal in consensus sequences from other genotypes. Why these particular transformations occur mainly in the anti-HBe positive patients is not clear. Similar transformations to amino acids characteristic for other genotypes were reported in another study, but little attention was paid to them (Carman et al, 1995). This can be a topic for speculation that perhaps the patients with these mutations are more likely to

develop the pre-core stop codon mutation or that the pre-core stop codon mutation is somehow associated with these mutations

Although many families were bled, only a few of the siblings were found to have the same amino acid substitution in the HBV core protein. Family clustering was only significant in patients with the pre-core stop codon mutation in a sense, that they shared this mutation. At the time of the study design, I expected that siblings will often carry the same amino acid substitution because possible similarity of immune pressure and HLA structure and it will be very easy to recognise members of different families by typical amino acid mutations in B and T cell epitopes. This was not found during the analysis of sequences, despite most of them were phylogenetically very close and children obviously got infected with the same viral strain.

In the surface gene, only one amino acid mutation 123 T to A was found, which could potentially affect the structure of the first loop of the a determinant of the HBsAg molecule, but it has never been associated with any clinical significance. The previously published vaccine escape mutation at HBsAg amino acid position 141 in The Gambia (Karthigeasu, 1994) was not observed in the chronic carriers studied here. This indicates that, if this mutation occurs, its prevalence is not high in the normal population. In four samples, HBsAg amino acid substitutions characteristic of other genotypes were observed. All these samples were anti-HBe positive. These observations are similar to the findings of this study in the HBV core gene and it is not clear why these substitutions occur and whether they have any clinical importance. Double infection with HBV genomes of two genotypes cannot be excluded, because only direct sequencing of PCR products was

performed. Cloning of these PCR products and subsequent sequencing would probably provide more information, why these variants occur.

Premature termination codons at position 216 of the S region of the HBV surface gene, preventing expression of full length HBsAg protein, previously described only in immunosuppressed patients (Preikschat et al, 1999), were found in three Gambian samples. There are many factors, which could cause immunosuppression in the Gambian population. Malnutrition, HIV infection and malaria infection are all known to facilitate immunodeficiency. The actual pathogenetic importance of this mutation is entirely unclear. These mutations did not change the amino acid sequence in the overlapping open reading frame of the polymerase gene. The expression of most important immunogenic parts of HBsAg protein cannot be influenced by this premature termination codon.

Nucleotide sequence comparisons have been used in the past not only to demonstrate transmission of the virus from mother to infant, but also to describe common source outbreaks of infection in different epidemiological settings. The presence of similar nucleotide sequences, or of well characterised variants of the virus, such as the pre-core stop codon and surface variants, or viruses with deletions, have been reliable indicators of transmission from one person to another (Hannoun et al, 1997; Lin et al, 1990; Bozkaya et al, 1996; McIntosh et al, 1998). Striking similarities of sequences of the pre-core/core genes were found in siblings and their mothers (Bozkaya et al, 1997). Some of the sequences remained identical over a period of twenty years. No data were available on families infected horizontally, or under other circumstances, or about HBV sequences in The Gambia. So, at the time of the study design it was very difficult to predict, what the findings would be. It was suggested previously that the children most likely get infected

from their siblings (Whittle et al, 1983) and it was decided to focus mainly on intrafamilial transmission, and to study gene sequences from two HBV genome fragments - the pre-core/core and surface genes. In the present study, I have attempted to address this question using rigorous phylogenetic analysis of sequences obtained from the surface and pre-core/core region of the virus, from family members and other chronic carriers.

Sequences of the surface gene fragments did not seem to provide the necessary information, since many of them were completely identical and therefore could only confirm that similar HBV strains were present in the families. However, phylogenetic analysis on surface gene was found to be an excellent tool for determining HBV genotypes, but not subtypes. Therefore one could speculate that from the evolutionary point of view distribution of HBV genomes into genotypes is more feasible than into subtypes.

Phylogenetic analysis of pre-core/core gene sequences provided more detailed information about the clustering of HBV infection in Gambian families and villages. Despite phylogenetic distances being very short, a clear evolutionary pattern emerged. Several clusters of viral strains were reported and some of them were supported with high bootstrap values (>70%).

At the geographical area level, it was not possible to see any significant clustering of pre-core/core sequences, because sequences from completely different villages and geographical areas were sometimes identical. This could suggest that extensive mixing of strains occurs at this level, most likely caused by the movement of infected individuals

among villages. If one could assume that HBV infection arrived in The Gambia relatively recently and spread from one part of the country to the rest, then it would be possible to see certain consistency in its distribution. However, the phylogenetic tree gives more impression that the infection has been present in this area for quite a long time and it is moving in different directions without significant sequence changes. This assumption would be against the theory, that HBV infection arrived relatively recently in Africa from the New World, suggested previously by phylogenetic analysis of different HBV strains (Bollyky et al, 1997 and 1999). Another explanation for this homogeneity of the HBV sequences could be that the number of possible mutations with similar survival capacity due to the overlapping reading frames is rather low and sequences get saturated very quickly. Reverse mutations may take place, thus complicating further evolutionary analysis of HBV genomes.

Both HBV genotypes A and E were clearly distinguishable in the phylogenetic trees. Thus, phylogenetic analysis can be used successfully for genotype detection not only on surface gene, but also on pre-core/core gene sequences. Distribution of both genotypes in the phylogenetic tree indicated that genotype A was introduced later and had less evolutionary changes (shorter evolutionary distance) or had some selection disadvantages, which could also explain its lower prevalence and limited spread within the country.

In some families pre-core/core gene sequences remained completely identical in the siblings over the period since their presumed infection. This is reminiscent of previous observations in the families, where children were known to be infected with HBV perinatally (Hannoun et al, 2000; Bozkaya et al, 1997). In addition, some sequences from the same Gambian village, but non-related subjects, remained completely identical for at

least ten year period, which as far as one could judge, has never been described previously. All patients with identical sequences were HBeAg positive. This is consistent with previous observations, that before seroconversion to anti-HBe mutations are rare (Bozkaya et al, 1996; Carman et al, 1995). The fact that children from different families, but from the same village carry the same viral sequence very strongly suggests that they get infected horizontally from a common source or from each other. Explanation for this surprising similarity of nucleotide sequences in non-related subjects after a long period of time could be the fact that population in the Gambian villages was relatively inbred and had very similar genetic backgrounds and therefore immune pressures were quite similar.

In many families and villages the sequences formed clusters, which were supported by high bootstrap value ($>70\%$), widely accepted as reliable criteria for the evaluation of the phylogenetic trees. Interestingly, some of the village and family clusters were very similar. For example, in the cluster from the village Kaiaf, sequences from families 79 and 80 were clustered together and these clusters were indistinguishable from each other, giving the impression that siblings from these families were infected from each other or from another common source in the village. Both families were unrelated, but inhabited the same compound.

16 HBV pre-core/core sequences from non related subjects in village Manduar were grouped in different phylogenetic clusters. These clusters were nearly indistinguishable by their size and phylogenetic distances from the family clusters in other phylogenetic trees. This could indicate that the routes of HBV transmission within families and within groups of children in the villages are similar. Interestingly, these clusters within the village were not related to compounds of birth. Most of the individuals were from different compounds

and still carried similar HBV genomes. Another less likely explanation could be that because of different genetic backgrounds these different groups within the village became carriers by being particularly susceptible to one specific strain of HBV out of the several strains circulating within the village.

Findings in other villages were similar to findings in Manduar, but only a small number of sequences were available from each village. Different small clusters from villages Baniakang, Kumbija, Damphakunda did not group together according to the village of birth, but were located in very different positions.

From these limited data we can assume that there is a clear distribution between different viruses within the village, but this distribution does not necessarily mean distribution according to families or compounds. It can be speculated that this distribution can be age related, but I cannot express a certain opinion on this matter as some age groups were under-represented in this study. To answer the question, more HBV sequences and relevant epidemiological or anthropological information has to be obtained from Manduar and other villages.

There were families, where HBV sequences definitely did not show any signs of clustering and were positioned all over the phylogenetic tree. By using genetic markers from host DNA, it was established that they were undoubtedly children from the same parents as their siblings. Unfortunately, information on whether the children grew up with their families, was not available. Traditionally, many families in The Gambia send their youngest children to their grandparents or other relatives in distant villages. That could certainly happen in early childhood or even during the weaning period, when the risk of

becoming a carrier is assumed to be the highest and therefore the possibility that these children with completely different sequences were infected in another place from the village of their birth could not be excluded. It is also very common, that children from one or even several compounds were kept together and handled by different mothers, increasing the risk of getting infected from other sources.

The importance of mother to child transmission in The Gambia was always uncertain from the previous studies (Whittle et al, 1983; Hall, personal communication). Studies in Manduar and Keneba showed no significant difference between the prevalence of HBsAg antigenemia in mothers of the children, who became chronic carriers compared with mothers of children, who did not become chronic carriers, but there was the trend (8 mothers of 35 chronic carriers and 3 mothers of 31 noncarriers were HBsAg- positive) (Whittle, 1990). In E- study, the prevalence of HBV carriage was significantly higher in children of HBeAg positive women (54%), than in HBeAg negative (26%) and non-carriers (19%) (Hall A, personal communication). This difference was not seen with HBsAg positive fathers. My study also showed higher prevalence (Vall et al, 1994) of the HBsAg antigenemia in the mothers of the carriers, than in the fathers. Prevalence of carriage in mothers was higher than average prevalence in Gambian women (10%). The phylogenetic analysis on HBV pre-core/core sequences in four mother and child pairs was performed. In three of them, mothers had very similar HBV sequences to their children, which clustered together in the phylogenetic tree. I consider this information in addition to epidemiological data about higher prevalence of HBsAg carriage in mothers sufficient to suggest that the mother to child transmission occurs in The Gambia, but it is definitely not the predominant route of transmission. When the actual mother to child transmission occurs is not clear. It was shown previously that in African population perinatal

transmission did not occur even in the HBeAg mothers (Marinier *et al*, 1985). Horizontal transmission or transmission via mothers milk most likely takes place.

A different statistical approach was designed in order to better characterise the clustering within villages and compounds, because bootstrap analysis could not provide enough statistical support for all clusters apparent by eye in the phylogenetic tree. Firstly, samples were divided according to their HBeAg status in positives and negatives, since it was shown that the HBeAg negatives had many more mutations over the time and this could diminish the accuracy of phylogenetic analysis. Secondly, both HBV genotypes A and E were analysed separately. In the genotype E samples, combined family + village tree, had much higher likelihood values, than either component tree, suggesting that these two levels of population structure contain most of the information. This statistical analysis supported the idea, that actual clustering is somewhere higher, than simple family clustering.

Therefore, based on epidemiological observations, phylogenetic trees and statistical analysis, a model of transmission of HBV in the Gambian population can be proposed. Children can get infected horizontally from their siblings, mothers, other compound members and even children outside the compounds, but that does not mean that all the children, even in the small villages, were infected from the same source or shared the same virus strain. Transmission seems to be multi-factorial and associated with the close contact between children. They could become infected through close contact when playing, sharing the same bed, or through other activities. How this transmission actually happens and what factors facilitate it, is still not clear and merits further anthropological investigation. The establishment of the single predominant risk factor is very unlikely.

8 Future plans

Epidemiology.

To go back and find more detailed epidemiological information about the siblings whose sequences were analysed in order to find out exactly where they were at the presumed time of the infection.

To bleed more people from the subjects of the same compounds and villages and vaccinated children of the families with multiple carriers in order to see whether they have got vaccine breakthrough infections.

To try to carry out a similar study to this in the country, where mass vaccination has not been introduced and all children in the families get exposed to infection in order to find the exact route of transmission by collecting anthropological and epidemiological information.

To use the similar combined approach to investigate an outbreak of the HBV and HCV infections in approximately fifty Latvian paediatric patients with acute leukemia who were infected during the chemotherapy in the ward over the period of two years.

Molecular biology

To obtain full genome HBV sequences from both Gambian genotypes A and E and try to use these sequences for the phylogenetic analysis of the evolution of the virus.

To clone some pre-core/core sequences in order to find quasispecies in different patients according to HBeAg status

To analyse further human genetic data from siblings or village inhabitants who carry related virus in order to see whether genetic background determines development of certain amino acid substitutions in the viral sequences.

Clinical interventions

To get more detailed information about the clinical features of the HBV infection in the carriers, involved in this study: e.g.liver biopsy in order to establish the activity of the disease.

To try to deliver the treatment currently available in The Gambia for those for whom it is necessary and to apply new treatments designed on the basis of this study.

Moral obligations

To keep an active follow up of the chronic HBV carriers identified during this study.

9 Appendices

Appendix I. Consent form

We are trying to find out why some people in The Gambia get infected with a virus (a very small infecting agent) called Hepatitis B, which then stays in their blood for a long time. This is a very common infection in The Gambia but only a few people carry the virus for a long time. We first need to test you, and your brothers and sisters, to see if you are carrying this virus. To do this, we need to take 2mls (half a teaspoon) of blood (show syringe) that will be taken back to MRC for testing. We will return to your compound to tell you the results. If you and one of your brothers or sisters are positive for the test, we will ask you for some more blood (10mls, two teaspoons), (show syringe) and will ask your parents if they will give the same amount of blood. These second specimens of blood will be carefully analysed to see if we can find out why you and your brother or sister has been carrying this infection. The knowledge gained from this research will help us understand more about this common infection in The Gambia and may help us to find a new treatment for this infection, which in some cases may cause damage to the liver.

Person's name_____

Field worker_____

Date __/__/__

Appendix II. Family tree questionnaire

	Surname	Name	Age	Vaccination	Sex
Father					
Mother					
Children					
1					
2					
3					

Appendix III. Field questionnaire

Name _____ Surname _____

Date of interview ____/____/____

Date of birth ____/____/____

Sex _____

Age in years _____

Village of birth _____

Village of current residence _____

Compound _____

Name of the compound owner _____

Name of the father _____ Surname _____

Name of the mother _____ Surname _____

Ethnic group

- | | |
|-------------|-------------|
| 1. Mandinka | 5. Fula |
| 2. Aku | 6. Sarahule |
| 3. Wollof | 7. Other |
| 4. Jola | |

Ethnic group of father?

- | | |
|-------------|-------------|
| 1. Mandinka | 5. Fula |
| 2. Aku | 6. Sarahule |
| 3. Wollof | 7. Other |
| 4. Jola | |

Ethnic group of mother?

- | | |
|-------------|-------------|
| 1. Mandinka | 5. Fula |
| 2. Aku | 6. Sarahule |
| 3. Wollof | 7. Other |
| 4. Jola | |

How many sisters and brothers index has
with the same father and mother? _____

How many brothers and sisters from the same parents
are older than the index? _____

What are family relationships with the index? _____

- | | |
|------------|-----------------------------|
| I- Index | C- Child |
| P- Parent | F- Half sibling from father |
| S- Sibling | M- Half sibling from mother |

First sample brady number _____

Second sample brady number _____

Date of first sample ____/____/____

Date of second sample ____/____/____

Directions to reach the house _____

Interviewer _____

FCNO _____

IDNO _____

PIDNO _____

DATEB ____/____/____

SEX _____

AGE _____

VILLB _____

VILLC _____

COMP _____

NCOMP _____

FATHN _____

MOTHN _____

ETNIC ____/____/____

ETNICF ____/____/____

ETNICM ____/____/____

FAMS ____/____/____

BORD ____/____/____

REL ____/____/____

SPNO 1 _____

SPNO2 _____

SPDATE1 ____/____/____

SPDATE2 ____/____/____

HBSAG ____/____/____

HBeAG ____/____/____

HBCAB ____/____/____

AST/____/____

ALT/____/____

Appendix 1V. Example of the result form

Dear parents,

Members of your family have been included in Chronic Hepatitis B and C family study. This study was designed to find out why some of your family members become carriers of hepatitis virus. First tests showed hepatitis B status of your family members.

Mariama Camara- HBsAg negative

Ebrima Camara- HBsAg positive

Saidou Camara- HBsAg negative

Amadou Camara- HBsAg negative

You may be invited to take part in other studies.

MRC, Fajara
Dr.Dumpis

Appendix V. Main serological markers and age from the family members of the previously known HBV chronic carriers

F	Father	Mother	Siblings in birth order							8,9sib
			1sib	2sib	3sib	4sib	5sib	6sib	7sib	
1	N	neg	+M17E	+F12	+M11	V				
2	N	N	N	+F18E	N	+M9E	V	V		
5	+	neg	M20	F14	+F12E	+M10E	V	V	V	
11	N	+	M19	+M17	F16	M14	+F11E	V	V	VV
12	N	+	+M17	F13	+F11E	V				
14	N	N	F31	N	F27	N	N	+F20	N	+F10
17	N	neg	N	+13F	-13M	+10M	V	V	V	
18	neg	N	+M20	N	+F14	+M11	+F9	+M7E	V	V V
20	neg	neg	+M15	+M13	N	+F10E				
21	neg	neg	+F13	N	+M10	V	V	V		
22	neg	neg	F20	N	N	+M11	+F9	V		
24	N	+	+M17	F15	+F10	V				
28	neg	neg	N	M20	N	+M15	+F12	+M12		
29	+	+	+F14	N	+M10	V				
30	N	N	+M14	+M10E						
31	N	neg	+M30	N	N	F22	M19	+M17E	+F12	+M10E
33	+	neg	+M16E	+M13E	+M10	+M8	V			
34	neg	neg	M18	+M15	+M13	+M10E				
35	neg	N	+M31	+M28	+M25E	N	N	+M12	N	
40	neg	N	N	F16	F14	+M12E	+F10E			
41	neg	+	+M18	+M16E	+M14	+M10				
42	N	neg	+M12E	+M10E						
44	N	neg	N	F19	+M16	M13	+F10			
47	neg	neg	M20	M14	+F12	+M10	V			
49	neg	neg	N	N	N	+M16	N	+F10	V	V, V
53	neg	+	+M16E	+M13E	+M11E					
57	neg	neg	+M18	N	N	+M13	F11	V	V	
59	N	neg	F26	M24	N	+F18	N	+M13	F10	
61	N	+	+F19E	+F17E	+M13E	+M11E	V			
69	neg	N	M20	+F17	F14	+F11	V	V	V	
72	neg	neg	F27	N	F23	F18	+M16E	N	+M12	
80	neg	neg	+M18E	F15	+F12E					
81	neg		+F14	+M11E	+M9E	N	N			
82	N	neg	+M12E	+F10E	V	V	V			
86	neg	neg	F30	F21	M19	N	+M15	M13	+M11	
93	neg	neg	F26	+M20	+M17	F12	+M10			
95	N	+	+F18	F12	+M10					
96	neg	neg	F21	F18	+F16	N	F12	+F10	V	V
97	N	neg	F15	+M13E	+F10	V	V	V		
99	N	neg	+M12	+M10E	V	V				
100	N	neg	+F30	+F28	+M26	+F21	M18	+F12		
101	neg	neg	F20	17M	+F15	F13	+M11	V	V	
106	N	neg	N	M20	F18	N	+M14	+M12	+F10	V V
108	neg	neg	M23	F21	F18	M15	+F15	+M10	V	
111	N	neg	23F	+20M	F15	M13	+F11	+F9		
114	N	neg	M19	+M16	+F14	V	V			
116	neg	neg	N	+M16E	+M13	M11	V	V	V	
117	neg	+E	F24	+M18	N	M15	+M13	V		
120	N	neg	N	+M13	+M11	V	V	V		
125	neg	neg	N	+M16	+M11	V	V			
126	+E	N	+F16	N	+F11					
127	N	+	+F12	+F10	V	V				
133	neg	neg	N	N	+F13	+M12E	V	V		
135	neg	N	F20	+M17	F15	F12	+M10E	V	V	
137	N	+	N	+M15E	+M12	+M10E	V	V	V	
139	N	+	N	+M21	+F17	N	F12	+F12	+F10	V
140	neg	neg	N	+F15	+M12					
141	+	neg	+M21	+F15E	+M13E					
142	neg	neg	+M21	+M14	V					
150	N	N	M50	M48	+F46	+F44				
153	+	neg	+M13E	+F11	+M10E					
154	neg	neg	+F19	+M16	M14	M12	N	V	V	V
158	N	+	+F12	+M10	V	V	V	V		
162	N	+	N	F20	+14M	M12	+M10	V	V	V V
165	N	neg	M15	+M13	+M10	V	V	V		
168	neg	N	F22	M19	+F13	+F10	V	V		
169	neg	neg	N	N	F21	N	+M16	+F12	V	
172	neg	+E	F13	+F11E	+F10E	V	V			
173	neg	neg	M28	+M21	+M21	F18	F18	+M14	+M10	
174	N	neg	+F24	N	+M19	N	+F14E	+F12	+F10E	V V
176	N	neg	F16	+M14	+M10	V				
181	N	neg	+F14	+M12	+F10E					
182	N	neg	M19	+M12E	+M10	V	V	V		
184	neg	+	+M13E	+M10	V	V	V	V		
187	N	neg	+M13	+M10E	V	V	V			
192	neg	+	F15	+M13	+F10	V	V			

199	neg	N	+M13	+M10	V	V	V			
202	neg	neg	N	N	N	+F13	+M10			
203	neg	+	F39	+M37	+F29	N	+F18	N	M13	
205	+	+	+F30	+M28	+F22	N	N			
207	N	+	+F38	+M28	M27	F24				
214	neg	neg	+F17	+F14E	+M11E	V	V	V	V	
217	neg	N	+F21	+F19E	+F15E	+F11E	V	V		
218	neg	N	+M25	N	+M20	N	N	N	+M11E	V
220	neg	neg	+M18	M16	+M11E	V				
222	neg	neg	+F23	F21	F18	M16	+M14	+F11	V	V
223	N	neg	+M20	+M15E	+M13	N	N	N		
225	neg	neg	M15	+M13E	+F11	V	V	V	V	
226	neg	neg	+F18	+M20	+M16E					
231	N	neg	+F29	N	+M24	+F22	+M19E	F17	+F15	
233	N	+	+F31	+F29						
234	neg	+	+F27	N	N	+M12				
237	N	N	+M24	+F23	N	N	N	N	N	
238	N	neg	F22	+M20E	+M16	+F14	V	V	V	
245	neg	neg	+M45	N	F40	+F38	F35	F32	N	N
246	N	N	+F36	F34	M28	+M23	N	N		
255	N	N	+F39	+F32	N	N	F22	N		
260	N	N	+M35	N	N	+M20	N	N		
261	N	N	+F50	N	N	M38	N	+F32		
263	N	N	+F28	+M25	N	+F21				
264	N	neg	+F36	N	+F26	N				
266	neg	neg	+M31	+F29	F23	+M20				
268	N	+	N	+F35	N	F23	+M23			
272	N	neg	M48	N	M38	N	+F30	+F27		
273	neg	neg	M30	+F27	N	+F22	F19	+M17	N	V
276	N	N	+F45	+F42	+F39	Not known				
277	N	+	N	N	+M13	+M14	+M15			

This table does not include families, where HBsAg status was already known for all family members or data on birth order were not available. Information about how many in the family were bled, how many were vaccinated and how many were not available at the time of the bleeding or refused to be bled is included in this picture. Age, sex, HBsAg, HBeAg status is also shown.

Abbreviations used in the table:

N - Not available at the time of bleeding or refused

+ - HBsAg positive, Neg - HBsAg negative, E - HBeAg positive

V- vaccinated

M- male, F- female

Numerical values - age

Appendix VI. Example of the FASTA sequence alignment format

Alignment of the fragments of surface genes.

```
>1003
AGGATCATCAACCACCAGTACGGGACCCTGCCGAACCTGCACGACTCTTGCTCAAGGA
ACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGACGGAAATTGCACTTGTATT
CCCATCCCATCATCATGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTT
CTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGCCGGGCTTTCCCCCATTGT
CTGGCTTTTAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGA
GTCCCTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGGG

>1073
AGGATCATCAACCACCAGTACGGGACCCTGTCTGAACCTGCACGACTCTTGCTCAAGGA
ACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGACGGAAATTGCACTTGTATT
CCCATCCCATCATCATGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTT
CTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGCCGGGCTTTCCCCCACTGT
CTGGCTTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGA
GTCCCTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGGG

>1112
AGGATCATCAACCACCAGTACGGGACCCTGCCGAACCTGCACGACTCTTGCTCAAGGA
ACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGACGGAAATTGCACTTGTATT
CCCATCCCATCATCATGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTT
CTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGCCGGGCTTTCCCCCACTGT
CTGGCTTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGA
GTCCCTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGGG
```

Appendix VII. Example of the NEXUS sequence alignment format

Three S gene fragments aligned.

```
#NEXUS

begin taxa;
  dimensions ntax=3;
  taxlabels
1003
1073
1112
;
end;

begin characters;
  dimensions nchar=338;
  format datatype=dna missing=N gap=-;
  matrix

1003
AGGATCATCAACCACCAGTACGGGACCCTGCCGAACCTGCACGACTCTTGCTCA
AGGAACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGAC
GGAAATTGCACTTGTAATCCCATCCCATCATCATGGGCTTTCGGAAAATTCCTAT
GGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGC
CATTTGTTTCAGTGGTTCGCCGGGCTTTCCTCCCATGCTGGCTTTTAGTTATATGG
ATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCC
CTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGCGG
1073
AGGATCATCAACCACCAGTACGGGACCCTGTGGAACCTGCACGACTCTTGCTCA
AGGAACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGAC
GGAAATTGCACTTGTAATCCCATCCCATCATCATGGGCTTTCGGAAAATTCCTAT
GGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGC
CATTTGTTTCAGTGGTTCGCCGGGCTTTCCTCCCATGCTGGCTTTTCAGTTATATGG
ATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCC
CTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGCGG
1112
AGGATCATCAACCACCAGTACGGGACCCTGCCGAACCTGCACGACTCTTGCTCA
AGGAACCTCTATGTTTCCCTCATGTTGCTGTTCAAAACCTTCGGAC
GGAAATTGCACTTGTAATCCCATCCCATCATCATGGGCTTTCGGAAAATTCCTAT
GGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGC
CATTTGTTTCAGTGGTTCGCCGGGCTTTCCTCCCATGCTGGCTTTTCAGTTATATGG
ATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCC
CTTTATACCTCTGTTACCAATTTTCTTTTGTCTCTGCGG

;
end;
begin assumptions;
  options deftype=unord;
end
```


Appendix VIII. Example of the log likelihood values for different simulated trees.

A genotype	E genotype	E genotype +	E genotype -
1328.12351	4038.09955	2202.85834	2131.48652
1333.4022	4057.57856	2238.27601	2140.69623
1336.97262	4080.44199	2268.61499	2144.87388
1337.2344	4088.87029	2282.8356	2146.43568
1339.89177	4091.42291	2285.38313	2147.89057
1341.37216	4091.95588	2286.43333	2150.09192
1344.0894	4093.08889	2294.32565	2153.06828
1345.10937	4096.42593	2302.09836	2153.32655
1346.28133	4096.48561	2306.17203	2153.45337
1348.19634	4098.51506	2307.21418	2154.05657
1349.84544	4100.65528	2311.09594	2154.25969
1350.13365	4101.4206	2314.9414	2154.94216
1350.20741	4103.95066	2315.04484	2155.09812
1352.01386	4105.33086	2317.17566	2155.77713
1352.31168	4105.48262	2318.33198	2156.05399
1353.79807	4106.32019	2320.21062	2156.39449
1354.23763	4106.97241	2320.74375	2158.91489
1354.79677	4107.73848	2321.84423	2159.52539
1356.31449	4108.72278	2322.31391	2159.74605
1356.52652	4110.27801	2323.25974	2160.31109
1356.8736	4110.2912	2323.46774	2160.60057

10 References

Akarca US, Greene S, Lok AS (1994) "Detection of precore hepatitis B virus mutants in asymptomatic HBsAg- positive family members", *Hepatology*, vol. 19,no. 6, pp. 1366-1370.

Akarca US, Lok AS (1995) "Naturally occurring hepatitis B virus core gene mutations", *Hepatology*, vol. 22,no. 1, pp. 50-60.

Alexopoulou A, Karayiannis P, Hadziyannis S J, Aiba N, Thomas HC. (1997) "Emergence and selection of HBV variants in an anti-HBe positive patient persistently infected with quasi-species", *Journal of Hepatology*, vol. 26,no. 4, pp. 748-753.

Allen JG, Tsuda F, Naito S, Takai E, Akahane Y, Furuta S, Miyakawa Y, Mayumi (1970) "Some factors regulating duration of incubation for serum hepatitis", *Archives of Surgery*, vol. 100,no. 1, pp. 2-5.

Almarri A, Batchelor JR (1994) "HLA and hepatitis B infection ", *Lancet*, vol. 344,no. 8931, pp. 1194-1195.

Almog Y, Klein A, Adler R, Laub O, Tur-Kaspa R (1992) "Estrogen suppresses hepatitis B virus expression in male athymic mice transplanted with HBV transfected Hep G-2 cells", *Antiviral Research.*, vol. 19,no. 4, pp. 285-293.

Alter HJ, Purcell RH, Gerin JL, London WT, Kaplan PM, McAuliffe V J, Wagner J, Holland PV.(1977) "Transmission of hepatitis B to chimpanzees by hepatitis B surface antigen-positive saliva and semen", *Infection and Immunity*, vol. 16,no. 3, pp. 928-933.

Anonymous.(1987). "The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group", *Cancer Research*, vol. 47,no. 21, pp. 5782-5787.

Aye TT, Bartholomeusz A, Shaw T, Bowden S, Breschkin A, McMillan J, Angus P, Locarnini S. (1997), "Hepatitis B virus polymerase mutations during antiviral therapy in a patient following liver transplantation", *Journal of Hepatology*, vol. 26,no. 5, pp. 1148-1153.

Aye TT, Uchida T, Becker SO, Shikata T, Mima S (1994) "Completely or nearly identical hepatitis B virus strains replicate between patients with acute or fulminant hepatitis B and their respective infectious sources", *Journal of Medical Virology*, vol. 42,no. 1, pp. 60-65.

Bah E, Hall A J, Inskip HM. (1990), "The first 2 years of the Gambian National Cancer Registry", *British Journal of Cancer*, vol. 62,no. 4, pp. 647-650.

Bancroft WH, Mundon FK, Russell PK. (1972), "Detection of additional antigenic determinants of hepatitis B antigen", *Journal of Immunology*, vol. 109,no. 4, pp. 842-848.

Bancroft WH, Snitbhan R, Scott RM, Tingpalapong M, Watson WT, Tanticharoenyos P, Karwacki J J, Srimarut S. (1977), "Transmission of hepatitis B virus to gibbons by exposure to human saliva containing hepatitis B surface antigen", *Journal of Infectious Diseases*, vol. 135,no. 1, pp. 79-85.

Baptista M, Kramvis A, Kew MC. (1999) "High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers", *Hepatology*, vol. 29,no. 3, pp. 946-953.

Barin F, Perrin J, Chotard J, Denis F, N'Doye R, Diop MI, Chiron JP, Coursaget P, Goudeau A, Maupas P. (1981), "Cross-sectional and longitudinal epidemiology of hepatitis B in Senegal", *Progress of Medical Virology*, vol. 27, pp. 148-162.

Barlet V, Zarski JP, Thelu MA, Bichard P, Seigneurin JM. (1994), "Different prevalence of precore mutants in five members of a hepatitis- B-virus-infected family: evidence for a precore variant type in an asymptomatic anti-HBs patient", *Journal of Hepatology*, vol. 21,no. 5, pp. 797-805.

Baumert TF, Marrone A, Vergalla J, Liang TJ. (1998), "Naturally occurring mutations define a novel function of the hepatitis B virus core promoter in core protein expression", *Journal of Virology*, vol. 72,no. 8, pp. 6785-6795.

Beasley RP, Trepo C, Stevens CE, Szmunness W. (1977), "The e antigen and vertical transmission of hepatitis B surface antigen", *American Journal of Epidemiology*, vol. 105,no. 2, pp. 94-98.

Beasley RP, Lin CC, Chien CS, Chen CJ, Hwang LY. (1982), "Geographic distribution of HBsAg carriers in China", *Hepatology*, vol. 2,no. 5, pp. 553-556.

Beasley RP, Hwang LY, Stevens CE, Lin CC, Hsieh FJ, Wang KY, Sun TS, Szmunness W. (1983), "Efficacy of hepatitis B immune globulin for prevention of perinatal transmission of the hepatitis B virus carrier state: final report of a randomized double-blind, placebo-controlled trial", *Hepatology*, vol. 3,no. 2, pp. 135-141.

Beasley RP. (1988), "Hepatitis B virus. The major etiology of hepatocellular carcinoma", *Cancer*, vol. 61,no. 10, pp. 1942-1956.

- Beasley RP, Hwang LY (1991). Overview on the epidemiology of hepatocellular carcinoma. *American Journal of Epidemiology*. vol. 85, pp 532-535.
- Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. (1998), "Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans ", *New England Journal of Medicine*, vol. 338,no. 10, pp. 640-644.
- Berk PD and Popper H (1978) "Fulminant hepatic failure". *American Journal of Epidemiology*. vol. 69, pp 349-400.
- Bernier RH, Sampliner R, Gerety R, Tabor E, Hamilton F, Nathanson N. (1982), "Hepatitis B infection in households of chronic carriers of hepatitis B surface antigen: factors associated with prevalence of infection", *American Journal of Epidemiology*, vol. 116,no. 2, pp. 199-211.
- Berris B, Wrobel DM, Sinclair JC, Feinman SV (1973), "Hepatitis B antigen in families of blood donors", *Annals of Internal Medicine*, vol. 79,no. 5, pp. 690-693.
- Bertoletti A, Ferrari C, Fiaccadori F, Penna A, Margolskee R, Schlicht HJ, Fowler P, Guilhot S, & Chisari FV. (1991), "HLA class I-restricted human cytotoxic T cells recognize endogenously synthesized hepatitis B virus nucleocapsid antigen", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88,no. 23, pp. 10445-10449.
- Bertoletti A, Costanzo A, Chisari FV, Levrero M, Artini M, Sette A, Penna A, Giuberti T, Fiaccadori F, Ferrari C. 1994, "Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope", *Journal of Experimental Medicine*, vol. 180,no. 3, pp. 933-943.

Bianchi L (1970), "Morphologic features in biopsy diagnosis of acute viral hepatitis", *Progress in Liver Diseases*, vol. 3, pp. 236-251.

Billewicz WZ, McGregor IA (1981). "The demography of West African (Gambian) villages". *Journal of Biosocial Sciences*, vol 13, pp.1951-1975.

Bodsworth NJ, Cooper DA, Donovan B. (1991), "The influence of human immunodeficiency virus type 1 infection on the development of the hepatitis B virus carrier state", *Journal of Infectious Diseases*, vol. 163,no. 5, pp. 1138-1140.

Bollyky PL, Rambaut A, Grassly N, Carman WP, Holmes EC. Hepatitis B virus has a recent new world evolutionary origin (2000). *Hepatology*, vol 26, pp.765-770.

Bollyky PL and Holmes EC. (1999), "Reconstructing the complex evolutionary history of hepatitis B virus", *Jornal of Molecular Evolution*, vol. 49,no. 1, pp. 130-141.

Bozkaya H, Ayola B, Lok AS. (1996), "High rate of mutations in the hepatitis B core gene during the immune clearance phase of chronic hepatitis B virus infection", *Hepatology*, vol. 24,no. 1, pp. 32-37.

Bozkaya H, Akarca US, Ayola B, Lok AS. (1997), "High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection", *Journal of Hepatology*, vol. 26,no. 3, pp. 508-516.

Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, Makris A, Thomas HC. (1989), "Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection", *Lancet*, vol. 2,no. 8663, pp. 588-591.

Carman WF, Zanetti AR, Karayiannis P, Waters J, Manzillo G, Tanzi E, Zuckerman A J, Thomas HC. (1990), "Vaccine-induced escape mutant of hepatitis B virus", *Lancet*, vol. 336,no. 8711, pp. 325-329.

Carman W, Thomas H, Domingo E (1993). "Viral genetic variation: hepatitis B virus as a clinical example". *The Lancet*, vol. 341, pp. 349-353.

Carman WF. (1997), "The clinical significance of surface antigen variants of hepatitis B virus.", *Journal of Viral Hepatitis*, vol. 4 Suppl 1, pp. 11-20.

Carman WF, Thursz M, Hadziyannis S, McIntyre G, Colman K, Gioustoz A, Fattovich G, Alberti A, Thomas HC. (1995), "Hepatitis B e antigen negative chronic active hepatitis: hepatitis B virus core mutations occur predominantly in known antigenic determinants", *Journal of Viral Hepatitis*, vol. 2,no. 2, pp. 77-84.

Carman WF, Boner W, Fattovich G, Colman K, Dornan ES, Thursz M, Hadziyannis S. (1997), "Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission", *Journal of Infectious Diseases*, vol. 175,no. 5, pp. 1093-1100.

Carotenuto P, Quinti I, Pontesilli O, Hall AJ, DeLange GG, Whittle HC, D'Amelio R, Aiuti F. (1995), "Response to hepatitis B vaccine in a cohort of Gambian children", *Pediatric Infectious Disease Journal*, vol. 14,no. 3, pp. 215-220.

Cavanaugh VJ, Guidotti LG, Chisari FV. (1997), "Interleukin-12 inhibits hepatitis B virus replication in transgenic mice", *Journal of Virology*, vol. 71,no. 4, pp. 3236-3243.

Chiou HL, Lee TS, Kuo J, Mau YC, Ho MS. (1997), "Altered antigenicity of 'a' determinant variants of hepatitis B virus", *Journal of General Virology*, vol. 78,no. Pt 10, pp. 2639-2645.

Chisari FV and Ferrari C. (1995), "Hepatitis B virus immunopathogenesis." *Annual Review of Immunology* , vol. 13, pp. 29-60.

Chisari FV and Ferrari C (1995), "Hepatitis B virus immunopathology.", *Springer Seminars in Immunopathology*, vol. 17,no. 2-3, pp. 261-281.

Chisari FV (1997), "Cytotoxic T cells and viral hepatitis." *Journal of Clinical Investigation*, vol. 99,no. 7, pp. 1472-1477.

Chong-Jin O, Wei NC, Shiuan K, Gek KL (1999), "Identification of hepatitis B surface antigen variants with alterations outside the "a" determinant in immunized Singapore infants", *Journal of Infectious Diseases*, vol. 179,no. 1, pp. 259-263.

Chuang WL, Chang WY, Lu SN, Lin ZY, Chen SC, Hsieh MY, Wang LY, You SL, Chen C J. (1993), "The role of hepatitis C virus in chronic hepatitis B virus infection", *Gastroenterology Japan*, vol. 28 Suppl 5, pp. 23-27.

Clewley JP. (1998), "A user's guide to producing and interpreting tree diagrams in taxonomy and phylogenetics. Part 4: Practice", *Communicable Disease and Public Health*, vol. 1,no. 4, pp. 285-287.

Cohen BJ. (1978), "The IgM antibody responses to the core antigen of hepatitis B virus", *Journal of Medical Virology*, vol. 3,no. 2, pp. 141-149.

Coleman JC, Waugh M, Dayton R. (1977), "Hepatitis B antigen and antibody in a male homosexual population", *British Journal of Venerological Diseases*, vol. 53,no. 2, pp. 132-134.

Coleman PF, Chen YC, Mushahwar IK. (1999), "Immunoassay detection of hepatitis B surface antigen mutants", *Journal of Medical Virology*, vol. 59,no. 1, pp. 19-24.

Corpet F. (1988), "Multiple sequence alignment with hierarchical clustering", *Nucleic Acids Research*, vol. 16,no. 22, pp. 10881-10890.

Coursaget P, Yvonnet B, Bourdil C, Mevelec MN, Adamowicz P, Barres JL, Chotard J, N'Doye R, Diop MI, Chiron JP. (1987), "HBsAG positive reactivity in man not due to hepatitis B virus", *Lancet*, vol. 2,no. 8572, pp. 1354-1358.

Craxi A, Montano L, Goodall A, Thomas HC. (1982), "Genetic and sex-linked factors influencing HBs antigen clearance. I. Nonimmune clearance in inbred strains of mice", *Journal of Medical Virology*, vol. 9,no. 2, pp. 117-123.

Daniels HM, Meager A, Eddleston AL, Alexander GJ, Williams R. (1990), "Spontaneous production of tumour necrosis factor alpha and interleukin-1 beta during interferon-alpha treatment of chronic HBV infection", *Lancet*, vol. 335,no. 8694, pp. 875-877.

Devi-Karthigesu V, Allison LMC, Fortuin M, Mendy ME, Whittle HC, Howard CR.(1994) A novel hepatitis B virus variant in the sera of immunised children. *Journal of General Virology*, vol. 75, pp. 443-448.

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J. (1996), "A comprehensive genetic map of the human genome based on 5,264 microsatellites", *Nature*, vol. 380,no. 6570, pp. 152-154.

Dienes HP, Gerken G, Goergen B, Heermann K, Gerlich W, Meyer BK. (1995), "Analysis of the precore DNA sequence and detection of precore antigen in liver specimens from patients with anti-hepatitis B e-positive chronic hepatitis", *Hepatology*, vol. 21,no. 1, pp. 1-7.

Dragosics B, Ferenci P, Hitchman E, Denk H. (1987), "Long-term follow-up study of asymptomatic HBsAg-positive voluntary blood donors in Austria: a clinical and histologic evaluation of 242 cases", *Hepatology*, vol. 7,no. 2, pp. 302-306.

Dutra SF, Fernandes FC, Coimbra GA.(1998), "Outbreak of hepatitis B virus in recent arrivals to the Brazilian Amazon", *Journal of Medical Virology*, vol. 56,no. 1, pp. 4-9.

Edmunds WJ, Medley GF, Nokes DJ, Hall AJ, Whittle HC. (1993), "The influence of age on the development of the hepatitis B carrier state", *Proceedings of the Royal Society of London - Series B: Biological Sciences*, vol. 253,no. 1337, pp. 197-201.

Edmunds WJ, Medley GF, Nokes DJ, O'Callaghan CJ, Whittle HC, Hall AJ. (1996), "Epidemiological patterns of hepatitis B virus in highly endemic areas", *Epidemiology & Infection* , vol. 117,no. 2, pp. 313-325.

Ehata T, Omata M, Yokosuka O, Hosoda K, Ohto M.(1992), "Variations in codons 84-101 in the core nucleotide sequence correlate with hepatocellular injury in chronic hepatitis B virus infection", *Journal of Clinical Investigation*, vol. 89,no. 1, pp. 332-338.

Encke J & Wands JR. (1999), "DNA vaccines", *Intervirology*, vol. 42,no. 2-3, pp. 117-124.

Fang ZL, Ling R., Wang SS, Nong J, Huang CS, Harrison TJ. (1998), "HBV core promoter mutations prevail in patients with hepatocellular carcinoma from Guangxi, China", *Journal of Medical Virology*, vol. 56,no. 1, pp. 18-24.

Felsenstein J. (1981), "Evolutionary trees from DNA sequences: a maximum likelihood approach", *Journal of Molecular Evolution*, vol. 17,no. 6, pp. 368-376.

Felsenstein J. (1988), "Phylogenies from molecular sequences: inference and reliability", *Annual Reviews in Genetics*, vol. 22, pp. 521-565.

Ferrari C, Penna A, Bertoletti A, Valli A, Antoni AD, Giuberti T, Cavalli A, Petit MA, Fiaccadori F. (1990), "Cellular immune response to hepatitis B virus-encoded antigens in acute and chronic hepatitis B virus infection", *Journal of Immunology*, vol. 145,no. 10, pp. 3442-3449.

Ferrari C, Bertoletti A, Penna A, Cavalli A, Valli A, Missale G, Pilli M, Fowler P, Giuberti T, Chisari FV. (1991), "Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen", *Journal of Clinical Investigation*, vol. 88,no. 1, pp. 214-222.

Fields BN (1997). *Fields virology*. Blackwell Science. Pp. 2748.

- Fortuin M, Chotard J, Jack AD, Maine NP, Mendy M, Hall AJ, Inskip HM, George MO, Whittle HC. (1993), "Efficacy of hepatitis B vaccine in the Gambian expanded programme on immunisation", *Lancet*, vol. 341,no. 8853, pp. 1129-1131.
- Fortuin M, Karthigesu V, Allison L, Howard C, Hoare S, Mendy M, Whittle HC. (1994), "Breakthrough infections and identification of a viral variant in Gambian children immunized with hepatitis B vaccine", *Journal of Infectious Diseases*, vol. 169,no. 6, pp. 1374-1376.
- Foster O, Ajdukiewicz A, Ryder R, Whittle H, Zuckerman AJ. (1984), "Hepatitis B virus transmission in West Africa: a role for tropical ulcer?", *Lancet*, vol. 1,no. 8376, pp. 576-577.
- Fujiwara K, Yokosuka O, Ehata T, Chuang WL, Imazeki F, Saisho H, Omata M. (1998), "The two different states of hepatitis B virus DNA in asymptomatic carriers: HBe-antigen-positive versus anti-HBe- positive asymptomatic carriers", *Digestive Diseases & Sciences*, vol. 43,no. 2, pp. 368-376.
- Fulford KW, Dane DS, Catterall RD, Woof R, Denning JV. (1973), "Australia antigen and antibody among patients attending a clinic for sexually transmitted diseases", *Lancet*, vol. 1,no. 7818, pp. 1470-1473.
- Gerlich WH. (1993). Structure and molecular biology. *Zuckerman AJ, Thomas HC. Viral hepatitis [ed 2]*, 89. Edinburgh, Churchill Livingstone.

Giani G, Chiaramonte M, Pasini CV, Fagiolo U & Naccarato R. (1979), "Hepatitis B surface antigenemia and HLA antigens ", *New England Journal of Medicine*, vol. 300,no. 18, p. 1056.

Gilles PN, Fey G, Chisari FV. (1992), "Tumor necrosis factor alpha negatively regulates hepatitis B virus gene expression in transgenic mice", *Journal of Virology*, vol. 66,no. 6, pp. 3955-3960.

Girones R, Miller RH. (1989), "Mutation rate of the hepadnavirus genome", *Virology*, vol. 170,no. 2, pp. 595-597.

Goldin RD, Fish DE, Hay A, Waters JA, McGarvey MJ, Main J, Thomas HC. (1990), "Histological and immunohistochemical study of hepatitis B virus in human immunodeficiency virus infection", *Journal of Clinical Pathology*, vol. 43,no. 3, pp. 203-205.

Grady GF & Lee VA. (1975), "Hepatitis B immune globulin--prevention of hepatitis from accidental exposure among medical personnel", *New England Journal of Medicine*, vol. 293,no. 21, pp. 1067-1070.

Greenfield C, Osidiana V, Karayiannis P, Galpin S, Musoke R, Jowett TP, Mati P, Tukei PM, Thomas HC. (1986), "Perinatal transmission of hepatitis B virus in Kenya: its relation to the presence of serum HBV-DNA and anti-HBe in the mother", *Journal of Medical Virology*, vol. 19,no. 2, pp. 135-142.

Guidotti LG, Chisari FV. (1999), "Cytokine-induced viral purging--role in viral pathogenesis", *Current Opinion in Microbiology*, vol. 2,no. 4, pp. 388-391.

- Guidotti LG, McClary H, Loudis JM, Chisari FV (2000). "Nitric oxide inhibits hepatitis B virus replication in the livers of transgenic mice", *Journal of Experimental Medicine*, vol. 191,no. 7, pp. 1247-1252.
- Guidotti LG, Rochford R, Chung J, Shapiro M, Purcell R, Chisari FV. (1999), "Viral clearance without destruction of infected cells during acute HBV infection", *Science*, vol. 284,no. 5415, pp. 825-829.
- Hadler SC, Judson FN, O'Malley PM, Altman NL, Penley K, Buchbinder S, Schable CA, Coleman PJ, Ostrow DN, Francis DP.(1991), "Outcome of hepatitis B virus infection in homosexual men and its relation to prior human immunodeficiency virus infection", *Journal of Infectious Diseases*, vol. 163,no. 3, pp. 454-459.
- Halliday ML, Kang LY, Rankin JG, Coates RA, Corey PN, Hu ZH, Zhou TK, Yuan GJ, Yao FL. (1992), "An efficacy trial of a mammalian cell-derived recombinant DNA hepatitis B vaccine in infants born to mothers positive for HBsAg, in Shanghai, China", *International Journal of Epidemiology*, vol. 21,no. 3, pp. 564-573.
- Hann HW, Kim CY, London WT, Whitford P, Blumberg BS. (1982), "Hepatitis B virus and primary hepatocellular carcinoma: family studies in Korea", *International Journal of Cancer*, vol. 30,no. 1, pp. 47-51.
- Hannoun C, Horal P, Lindh M. (1999), "Long-term mutation rates in the hepatitis B virus genome", *Journal of General Virology*, vol. 81 Pt 1, pp. 75-83.
- Hawkins AE, Zuckerman MA, Briggs M, Gilson RJ, Goldstone AH, Brink NS, Tedder RS. "Hepatitis B nucleotide sequence analysis: linking an outbreak of acute hepatitis B to

contamination of a cryopreservation tank". *Journal of Virological Methods* 1996 60(1):81-

8

Henigst W (1973), "Letter: Sexual transmission of infections associated with hepatitis-B antigen", *Lancet*, vol. 2,no. 7842, p. 1395.

Ho MS, Lu CF, Kuo J, Mau YC, Chao WH. (1995), "A family cluster of an immune escape variant of hepatitis B virus infecting a mother and her two fully immunized children", *Clinical and Diagnostic Laboratory Immunology*, vol. 2,no. 6, pp. 760-762.

Hohler T, Gerken G, Notghi A, Lubjuhn R, Taheri H, Protzer U, Lohr HF, Schneider PM., Meyer zum Buschenfelde KH, Rittner C. (1997), "HLA-DRB1*1301 and *1302 protect against chronic hepatitis B", *Journal of Hepatology*, vol. 26,no. 3, pp. 503-507.

Holmes EC, Zhang LQ, Simmonds P, Ludlam CA, Brown AJ. (1992), "Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient", *Proceedings of National Academy of Sciences U.S.A.*, vol. 89,no. 11, pp. 4835-4839.

Hou J, Karayiannis P, Waters J, Luo K, Liang C, Thomas HC. (1995), "A unique insertion in the S gene of surface antigen--negative hepatitis B virus Chinese carriers", *Hepatology*, vol. 21,no. 2, pp. 273-278.

Hsieh CC, Tzonou A, Zavitsanos X, Kaklamani E, Lan SJ, Trichopoulos D. (1992) Age at first establishment of chronic hepatitis B virus infection and hepatocellular carcinoma risk. A birth order study. *American Journal of Epidemiology*, **136**, 1115-1121.

Hsu HY, Chang MH, Liaw SH, Ni YH , Chen HL. (1999), "Changes of hepatitis B surface antigen variants in carrier children before and after universal vaccination in Taiwan". *Hepatology* 30, pp 1312-1317.

Huelsenbeck JP, Rannala B. (1997), "Phylogenetic methods come of age: testing hypotheses in an evolutionary context ", *Science*, vol. 276,no. 5310, pp. 227-232.

Inaba N, Ohkawa R, Matsuura A, Kudoh J, Takamizawa H. (1979), "Sexual transmission of hepatitis B surface antigen. Infection of husbands by HBsAg carrier-state wives", *British Journal of Veneral Diseases*, vol. 55,no. 5, pp. 366-368.

Inskip HM, Hall AJ, Temple IK, Loik F, Herbage E, Chotard J, Whittle H. (1991), "Response to hepatitis B vaccine in relation to the hepatitis B status of family members", *International Journal of Epidemiology*, vol. 20,no. 3, pp. 770-773.

Jacobs B, Mayaud P, Chagalucha J, Todd J, Ka-Gina G, Grosskurth H, Berege ZA. (1997), "Sexual transmission of hepatitis B in Mwanza, Tanzania", *Sexually Transmitted Diseases*, vol. 24,no. 3, pp. 121-126.

Jung MC, Diepolder HM, Pape GR. (1994), "T cell recognition of hepatitis B and C viral antigens.", *European Journal of Clinical Investigation*, vol. 24,no. 10, pp. 641-650.

Jung MC, Diepolder HM, Spengler U, Wierenga EA, Zachoval R, Hoffmann RM, Eichenlaub D, Frosner G, Will H , Pape GR. (1995), "Activation of a heterogeneous hepatitis B (HB) core and e antigen-specific CD4+ T-cell population during seroconversion to anti-HBe and anti-HBs in hepatitis B virus infection", *Journal of Virology*, vol. 69,no. 6, pp. 3358-3368.

Kaneko M, Uchida T, Moriyama M, Arakawa Y, Shikata T, Gotoh K, Mima S. (1995), "Probable implication of mutations of the X open reading frame in the onset of fulminant hepatitis B", *Journal of Medical Virology*, vol. 47,no. 3, pp. 204-208.

Karasawa T, Shirasawa T, Okawa Y, Kuramoto A, Shimada N, Aizawa Y, Zeniya M , Toda G. (1997), "Association between frequency of amino acid changes in core region of hepatitis B virus (HBV) and the presence of precore mutation in Japanese HBV carriers", *Journal of Gastroenterology*, vol. 32,no. 5, pp. 611-622.

Karayiannis P, Alexopoulou A, Hadziyannis S, Thursz M, Watts R, Seito S, Thomas HC. (1995), "Fulminant hepatitis associated with hepatitis B virus e antigen- negative infection: importance of host factors", *Hepatology*, vol. 22,no. 6, pp. 1628-1634.

Karthigesu VD, Allison LM, Fortuin M, Mendy M, Whittle HC, Howard CR. (1994), "A novel hepatitis B virus variant in the sera of immunized children", *Journal of General Virology*, vol. 75,no. Pt 2, pp. 443-448.

Kekule AS, Lauer U, Meyer M, Caselmann WH, Hofschneider PH, Koshy R. (1990), "The preS2/S region of integrated hepatitis B virus DNA encodes a transcriptional transactivator", *Nature*, vol. 343,no. 6257, pp. 457-461.

Kidd-Ljunggren K, Oberg M, Kidd AH. (1995), "The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences", *Journal of General Virology*, vol. 76,no. Pt 9, pp. 2119-2130.

Knutsson M, Kidd-Ljunggren K (2000). "Urine from chronic hepatitis B virus carriers: implications for infectivity", *Journal of Medical Virology* vol. 60,no. 1, pp. 17-20.

Kramvis A, Bukofzer S, Kew MC, Song E. (1997), "Nucleic acid sequence analysis of the precore region of hepatitis B virus from sera of southern African black adult carriers of the virus", *Hepatology*, vol. 25,no. 1, pp. 235-240.

Laskus T, Rakela J, Nowicki MJ, Persing DH. (1995), "Hepatitis B virus core promoter sequence analysis in fulminant and chronic hepatitis B", *Gastroenterology*, vol. 109,no. 5, pp. 1618-1623.

Laskus T, Rakela J, Persing DH (1995), "Nucleotide sequence analysis of precore and proximal core regions in patients with chronic hepatitis B treated with interferon", *Digestive Diseases Science*, vol. 40,no. 1, pp. 1-7.

Le Bouvier GL (1971), "The heterogeneity of Australia antigen", *Journal of Infectious Diseases*, vol. 123,no. 6, pp. 671-675.

Lee WM. (1997), "Hepatitis B virus infection ", *New England Journal of Medicine*, vol. 337,no. 24, pp. 1733-1745.

Lemon SM, Gates NL, Simms TE, Bancroft WH. (1981), "IgM antibody to hepatitis B core antigen as a diagnostic parameter of acute infection with hepatitis B virus", *Journal of Infectious Diseases*, vol. 143,no. 6, pp. 803-809.

Liang TJ, Hasegawa K, Rimon N, Wands JR, Ben-Porath E. (1991), "A hepatitis B virus mutant associated with an epidemic of fulminant hepatitis", *New England Journal of Medicine*, vol. 324,no. 24, pp. 1705-1709.

Lin TM, Chen CJ, Wu MM, Yang CS, Chen JS, Lin CC, Kwang TY, Hsu ST, Lin SY, Hsu LC. (1989), "Hepatitis B virus markers in Chinese twins", *Anticancer Research*, vol. 9,no. 3, pp. 737-741.

Lin HJ, Lai CL, Lau JY, Chung HT, Lauder IJ, Fong MW. (1990), "Evidence for intrafamilial transmission of hepatitis B virus from sequence analysis of mutant HBV DNAs in two Chinese families", *Lancet*, vol. 336,no. 8709, pp. 208-212.

Lindh M, Horal P, Dhillon AP, Furuta Y, Norkrans G. (1996), "Hepatitis B virus carriers without precore mutations in hepatitis B e antigen-negative stage show more severe liver damage", *Hepatology*, vol. 24,no. 3, pp. 494-501.

Lindh M, Andersson AS, Gusdal A. (1997), "Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus; large-scale analysis using a new genotyping method", *Journal of Infectious Diseases*, vol. 175,no. 6, pp. 1285-1293.

London WT, Drew JS, O'Grady JG, Alexander GJ, Hayllar KM, Williams R. (1977), "Sex differences in response to hepatitis B infection among patients receiving chronic dialysis treatment", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74,no. 6, pp. 2561-2563.

Magnius L, Kaplan L, Vyas GN, Perkins HA. (1975), "A new virus specified determinant of hepatitis B surface antigen", *Acta Pathologica et Microbiologica Scandinavia*, vol. 83,no. 3, pp. 295-297.

Maini MK, Boni C, Ogg GS, King AS, Reignat S, Lee CK, Larrubia JR., Webster GJ, McMichael AJ, Ferrari C, Williams R, Vergani D, Bertoletti A. (1999), "Direct ex vivo analysis of hepatitis B virus-specific CD8(+) T cells associated with the control of infection", *Gastroenterology*, vol. 117,no. 6, pp. 1386-1396.

Maini MK, Boni C, Lee CK, Larrubia JR, Reignat S, Ogg GS, King AS, Herberg J, Gilson R, Alisa A, Williams R, Vergani D, Naoumov NV, Ferrari C and Bertoletti A.(2000) The

role of virus-specific CD8(+) cells in liver damage and viral control during persistent hepatitis B virus infection. *Journal of Experimental Medicine* vol. Apr 17;191(8): pp. 1269-80

Marinier E, Barrois V, Larouze B, London WT, Cofer A, Diakhate L, Blumberg BS. (1985), "Lack of perinatal transmission of hepatitis B virus infection in Senegal, West Africa", *Journal of Pediatrics*, vol. 106,no. 5, pp. 843-849.

Marinos G, Torre F, Gunther S, Thomas MG, Will H, Williams R, Naoumov NV. (1996), "Hepatitis B virus variants with core gene deletions in the evolution of chronic hepatitis B infection", *Gastroenterology*, vol. 111,no. 1, pp. 183-192.

Martinson FE, Weigle KA, Royce RA, Weber DJ, Suchindran CM, Lemon SM. (1998), "Risk factors for horizontal transmission of hepatitis B virus in a rural district in Ghana", *American Journal of Epidemiology*, vol. 147,no. 5, pp. 478-487.

McClary H, Koch R, Chisari FV, Guidotti LG.(2000) "Relative sensitivity of hepatitis B virus and other hepatotropic viruses to the antiviral effects of cytokines ", *Journal of Virology*, vol. 74,no. 5, pp. 2255-2264.

McGregor IA. (1976). Health and communicable disease in a rural African environment. *Oikos*, pp. 186-192. .

McIntosh ED, Givney R, Zhang SS, Courouce AM, Burgess M, Cossart YE. (1998), "Molecular epidemiology and variation of hepatitis B in recent immigrant families to Australia", *Journal of Medical Virology*, vol. 56,no. 1, pp. 10-17.

McMahon BJ, Alward WL, Hall DB, Heyward WL, Bender TR, Francis DP, Maynard J E. (1985), "Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state", *Journal of Infectious Diseases*, vol. 151,no. 4, pp. 599-603.

McMahon BJ, Alberts SR, Wainwright RB, Bulkow L, Lanier AP. (1990), "Hepatitis B-related sequelae. Prospective study in 1400 hepatitis B surface antigen-positive Alaska native carriers", *Archives of Internal Medicine*, vol. 150,no. 5, pp. 1051-1054.

McMahon G, Ehrlich PH, Moustafa ZA, McCarthy LA, Dottavio D, Tolpin MD, Nadler PI, Ostberg L. (1992), "Genetic alterations in the gene encoding the major HBsAg: DNA and immunological analysis of recurrent HBsAg derived from monoclonal antibody-treated liver transplant patients", *Hepatology*, vol. 15,no. 5, pp. 757-766.

Mizokami M, Orito E, Ohba K, Ikeo K, Lau JY, Gojobori T. (1997), "Constrained evolution with respect to gene overlap of hepatitis B virus", *Journal of Molecular Evolution*, vol. 44 Suppl 1, p. S83-S90.

Mosley JW. (1975), "The epidemiology of viral hepatitis: an overview.", *American Journal of the Medical Sciences*, vol. 270,no. 2, pp. 253-270.

Mosley JW, Edwards VM, Casey G, Redeker AG, White E. (1975), "Hepatitis B virus infection in dentists", *New England Journal of Medicine*, vol. 293,no. 15, pp. 729-734.

Mota AH, Fainboim H, Terg R, Fainboim L. (1987), "Association of chronic active hepatitis and HLA B35 in patients with hepatitis B virus", *Tissue Antigens*, vol. 30,no. 5, pp. 238-240.

Munoz LA (1987). Serum opsonic activity and polymorphonuclear cell function in patients with chronic liver diseases. Ph thesis . University of London.

Nagasaka A, Hige S, Marutani M, Tsunematsu I, Saito M, Yamamoto Y, Konishi S, Asaka M. (1998), "Prevalence of mutations in core promoter/precore region in Japanese patients with chronic hepatitis B virus infection", *Digestive Disease Science*, vol. 43,no. 11, pp. 2473-2478.

Nassal M. "Hepatitis B virus replication: novel roles for virus-host interactions". *Intervirology*, vol 42,pp. 100-116

Nielsen JO, Dietrichson O, Elling P, Christoffersen, P. (1971), "Incidence and meaning of persistence of Australia antigen in patients with acute viral hepatitis: development of chronic hepatitis", *New England Journal of Medicine*, vol. 285,no. 21, pp. 1157-1160.

Norder H, Hammas B, Lofdahl S, Courouce AM, Magnus LO. (1992), "Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains", *Journal of General Virology.*, vol. 73,no. Pt 5, pp. 1201-1208.

Norder H, Courouce AM, Magnus LO (1993). "Complete nucleotide sequences of six hepatitis B viral genomes encoding the surface antigen subtypes ayw4, adw4q and adrq- and their phylogenetic classification". *Archives of Virology* suppl[8], 189-199.

Norder H, Courouce AM, Magnius LO. (1994), "Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes", *Virology*, vol. 198,no. 2, pp. 489-503.

Novacek MJ. (1992), "Mammalian phylogeny: shaking the tree", *Nature*, vol. 356,no. 6365, pp. 121-125.

O'Grady JG, Alexander GJ, Hayllar KM, Williams R. (1989), "Early indicators of prognosis in fulminant hepatic failure", *Gastroenterology*, vol. 97,no. 2, pp. 439-445.

Ogston CW, Wittenstein S., London WT, Millman I. (1979), "Persistence of hepatitis B surface antigen in the bedbug *Cimex hemipterus* (Fabr.)", *Journal of Infectious Diseases*, vol. 140,no. 3, pp. 411-414.

Ogston CW, London WT. (1980), "Excretion of hepatitis B surface antigen by the bedbug *Cimex hemipterus* Fabr", *Transactions of Royal Society of Tropical Medicine and Hygiene*, vol. 74,no. 6, pp. 823-825.

Okamoto H, Imai M, Shimozaki M, Hoshi Y, Iizuka H, Gotanda T, Tsuda F, Miyakawa Y, Mayumi M. (1986), "Nucleotide sequence of a cloned hepatitis B virus genome, subtype ayr: comparison with genomes of the other three subtypes", *Journal of General Virology*, vol. 67,no. Pt 11, pp. 2305-2314.

Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi, M. (1988), "Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes", *Journal of General Virology*, vol. 69,no. Pt 10, pp. 2575-2583.

Oldstone MB. (1991), "Molecular anatomy of viral persistence", *Journal of Virology*, vol. 65,no. 12, pp. 6381-6386.

Orito E, Mizokami M, Ina Y, Moriyama EN, Kameshima N, Yamamoto M, Gojobori T. (1989), "Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences", *Proceedings of National Academy of Sciences U.S.A.*, vol. 86,no. 18, pp. 7059-7062.

Oshima A, Tsukuma H, Hiyama T, Fujimoto I, Yamano H, Tanaka M. (1984), "Follow-up study of HBs Ag-positive blood donors with special reference to effect of drinking and smoking on development of liver cancer", *International Journal of Cancer*, vol. 34,no. 6, pp. 775-779.

Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JI, Schochetman G, Berkelman RL, Economou AN. (1992), "Molecular epidemiology of HIV transmission in a dental practice", *Science*, vol. 256,no. 5060, pp. 1165-1171.

Penna A, Artini M, Cavalli A, Levrero M, Bertoletti A, Pilli M, Chisari FV, Rehermann B, Del Prete G, Fiaccador F, Ferrari C. (1996), "Long-lasting memory T cell responses following self-limited acute hepatitis B", *Journal of Clinical Investigation*, vol. 98,no. 5, pp. 1185-1194.

Pisani P, Parkin DM, Munoz N, Ferlay J. (1997), "Cancer and infection: estimates of the attributable fraction in 1990", *Cancer Epidemiology and Biomarkers Preview*, vol. 6,no. 6, pp. 387-400.

Pollicino T, Campo S, Raimondo G. (1995), "PreS and core gene heterogeneity in hepatitis B virus (HBV) genomes isolated from patients with long-lasting HBV chronic infection", *Virology*, vol. 208,no. 2, pp. 672-677.

Pollicino T, Zanetti AR, Cacciola I, Petit MA, Smedile A, Campo S, Sagliocca L, Pasquali M, Tanzi E, Longo G, Raimondo G. (1997), "Pre-S2 defective hepatitis B virus infection in patients with fulminant hepatitis", *Hepatology*, vol. 26,no. 2, pp. 495-499.

Popper H, Roth L, Purcell RH, Tennant BC, Gerin JL. (1987), "Hepatocarcinogenicity of the woodchuck hepatitis virus", *Proceedings of National Academy of Sciences U.S.A.*, vol. 84,no. 3, pp. 866-870.

Power JP, Lawlor E, Davidson F, Holmes EC, Yap PL, Simmonds P. (1995), "Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin", *Lancet*, vol. 345,no. 8959, pp. 1211-1213.

Preikschat P, Meisel H, Will H, Gunther S. (1999), "Hepatitis B virus genomes from long-term immunosuppressed virus carriers are modified by specific mutations in several regions", *Journal of General Virology*, vol. 80,no. Pt 10, pp. 2685-2691.

Realdi G, Alberti A, Rugge M, Bortolotti F, Rigoli AM, Tremolada F, Ruol A. (1980), "Seroconversion from hepatitis B e antigen to anti-HBe in chronic hepatitis B virus infection", *Gastroenterology*, vol. 79,no. 2, pp. 195-199.

Realdi G, Fattovich G, Pastore G, Caredda F, Noventa F, Santantonio T, Moroni M, Criscuolo D, Maladorno D, Rugge M. (1990), "Problems in the management of chronic hepatitis B with interferon: experience in a randomized, multicentre study", *Journal of Hepatology*, vol. 11 Suppl 1, pp. S129-32.

Rehermann B, Fowler P, Sidney J, Person J, Redeker A, Brown M, Moss B, Sette A, Chisari FV. (1995), "The cytotoxic T lymphocyte response to multiple hepatitis B virus

- polymerase epitopes during and after acute viral hepatitis", *Journal of Experimental Medicine*, vol. 181,no. 3, pp. 1047-1058.
- Rimland D, Parkin WE, Miller GBJ, Schrack WD (1977), "Hepatitis B outbreak traced to an oral surgeon", *New England Journal of Medicine*, vol. 296,no. 17, pp. 953-958.
- Rodriguez-Frias F, Buti M, Jardi R, Cotrina M, Viladomiu L, Esteban R, Guardia J.(1995), "Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations", *Hepatology*, vol. 22,no. 6, pp. 1641-1647.
- Ross RK, Yu MC, Henderson BE, Yuan JM, Qian GS, Tu JT, Gao YT, Wogan GN, Groopman JD. (1992), "Aflatoxin biomarkers", *Lancet*, vol. 340,no. 8811, p. 119.
- Rossner MT. (1992), "Review: hepatitis B virus X-gene product: a promiscuous transcriptional activator", *Journal of Medical Virology*, vol. 36,no. 2, pp. 101-117.
- Ryder RW, Whittle HC, Wojciecowsky T, Moffat WM, Baker BA, Sarr E, Oldfield F. (1984) Screening for hepatitis B virus markers is not justified in West African transfusion centres. *Lancet* , vol. 69, pp. 349-400.
- Ryder RW, Whittle HC, Sanneh AB, Ajdukiewicz AB, Tulloch S, Yvonnet B. (1992), "Persistent hepatitis B virus infection and hepatoma in The Gambia, west Africa. A case-control study of 140 adults and their 603 family contacts", *American Journal of Epidemiology*, vol. 136,no. 9, pp. 1122-1131.

Sanderson MJ (1997). Objections to bootstrapping phylogenies: a critique. *Systematic Biology* vol. 44, pp. 299-320.

Sandler SG, Courouce AM, Soulier JP. (1978), "wVariants of the hepatitis B surface antigen in Israeli populations", *Vox Sanguinis*, vol. 34,no. 3, pp. 149-151.

Santantonio T, Jung MC, Pastore G, Angarano G, Gunther S, Will H. (1997), "Familial clustering of HBV pre-C and pre-S mutants", *Journal of Hepatology*, vol. 26,no. 2, pp. 221-227.

Scheuer PJ, Davies SE, Dhillon AP. (1996), "Histopathological aspects of viral hepatitis", *Journal of Viral Hepatitis*, vol. 3,no. 6, pp. 277-283.

Schmilovitz-Weiss H, Levy M, Thompson N, Dusheiko G. (1993), "Viral markers in the treatment of hepatitis B and C", *Gut*, vol. 34,no. 2 Suppl, p. S26-S35.

Schreeder MT, Thompson SE, Hadler SC, Berquist KR, Zaidi A, Maynard JE, Ostrow D, Judson FN, Braff EH, Nylund T, Moore JNJ, Gardner P, Doto IL, Reynolds G. (1982), "Hepatitis B in homosexual men: prevalence of infection and factors related to transmission", *Journal of Infectious Diseases*, vol. 146,no. 1, pp. 7-15.

Schweitzer IL, Mosley JW, Ashcaval M, Edwards VM, Overby LB. (1973), "Factors influencing neonatal infection by hepatitis B virus", *Gastroenterology*, vol. 65,no. 2, pp. 277-283.

Scully LJ, Brown D, Lloyd C, Shein R, Thomas HC. (1990), "Immunological studies before and during interferon therapy in chronic HBV infection: identification of factors predicting response", *Hepatology*, vol. 12,no. 5, pp. 1111-1117.

Seddigh-Tonekaboni S, Waters JA, Jeffers S, Gehrke R, Ofenloch B, Horsch A, Hess G, Thomas HC, Karayiannis P. (2000) "Effect of variation in the common "a" determinant on the antigenicity of hepatitis B surface antigen", *Journal of Medical Virology*, vol. 60,no. 2, pp. 113-121.

Seeff LB, Beebe GW, Hoofnagle JH, Norman JE, Buskell-Bales Z, Waggoner JG, Kaplowitz N, Koff RS, Petrini JLJ, Schiff ER. (1987), "A serologic follow-up of the 1942 epidemic of post-vaccination hepatitis in the United States Army", *New England Journal of Medicine*, vol. 316,no. 16, pp. 965-970.

Shiina S, Fujino H, Kawabe T, Tagawa K, Unuma T, Yoneyama M, Ohmori T, Suzuki S, Kurita M., Ohashi Y. (1991), "Relationship of HBsAg subtypes with HBeAg/anti-HBe status and chronic liver disease. Part II: Evaluation of epidemiological factors and suspected risk factors of liver dysfunction", *American Journal of Gastroenterology*, vol. 86,no. 7, pp. 872-875.

Sjogren MH and Lemon SM. (1983), "Low-molecular-weight IgM antibody to hepatitis B core antigen in chronic infections with hepatitis B virus", *Journal of Infectious Diseases*, vol. 148,no. 3, pp. 445-451.

Sterneck M, Kalinina T, Gunther S, Fischer L, Santantonio T, Greten H, Will H. (1998), "Functional analysis of HBV genomes from patients with fulminant hepatitis", *Hepatology*, vol. 28,no. 5, pp. 1390-1397.

Sterneck M, Kalinina T, Otto S, Gunther S, Fischer L, Burdelski M, Greten H, Broelsch CE, Will H. (1998), "Neonatal fulminant hepatitis B: structural and functional analysis of complete hepatitis B virus genomes from mother and infant", *Journal of Infectious Diseases*, vol. 177,no. 5, pp. 1378-1381.

Stevens CE, Taylor PE, Tong MJ, Toy PT, Vyas GN, Nair PV, Weissman JY, Krugman S. (1987), "Yeast-recombinant hepatitis B vaccine. Efficacy with hepatitis B immune globulin in prevention of perinatal hepatitis B virus transmission", *JAMA*, vol. 257,no. 19, pp. 2612-2616.

Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, Rossau R.(2000) "A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness", *Journal of General Virology*, vol. 81 Pt 1, pp. 67-74.

Szmunness W, Much I, Prince AM, Hoofnagle JH, Cherubin CE, Harley EJ, Block GH. (1975), "On the role of sexual behavior in the spread of hepatitis B infection", *Annals of Internal Medicine*, vol. 83,no. 4, pp. 489-495.

Szmunness W, Stevens CE, Harley EJ, Zang EA, Oleszko WR, William DC, Sadovsky R, Morrison JM, Kellner A. (1980), "Hepatitis B vaccine: demonstration of efficacy in a controlled clinical trial in a high-risk population in the United States", *New England Journal of Medicine*, vol. 303,no. 15, pp. 833-841.

Szmunness W, Stevens CE, Ikram H, Much MI, Harley EJ, Hollinger B. (1978), "Prevalence of hepatitis B virus infection and hepatocellular carcinoma in Chinese-Americans", *Journal of Infectious Diseases*, vol. 137,no. 6, pp. 822-829.

Tabor E, Bayley AC, Cairns J, Pelleu L, Gerety RJ. (1985), "Horizontal transmission of hepatitis B virus among children and adults in five rural villages in Zambia", *Journal of Medical Virology*, vol. 15,no. 2, pp. 113-120.

Tassopoulos NC, Papaevangelou GJ, Sjogren MH, Roumeliotou-Karayannis A, Gerin J. L, Purcell, R. H. (1987), "Natural history of acute hepatitis B surface antigen-positive hepatitis in Greek adults", *Gastroenterology*, vol. 92,no. 6, pp. 1844-1850.

ter Borg F, Jones EA. (1998), "Prediction of hepatic inflammatory activity in hepatitis B ", *Lancet*, vol. 352,no. 9139, p. 1555.

Thio CL, Thomas DL, Carrington M.(2000) "Chronic viral hepatitis and the human genome", vol. 31,no. 4, pp. 819-827.

Thomas HC, Foster GR, Sumiya M, McIntosh D, Jack DL, Turner MW, Summerfield J. A. (1996), "Mutation of gene of mannose-binding protein associated with chronic hepatitis B viral infection ", *Lancet*, vol. 348,no. 9039, pp. 1417-1419.

Thursz M, Kwiatkowski D, McGuire. (1996), "TNF promoter polymorphism and the outcome of hepatitis B virus infection". *Gut* vol. 39(suppl 1):A19. 1996.

Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, Hill AV. (1995), "Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia", *New England Journal of Medicine*, vol. 332,no. 16, pp. 1065-1069.

Tiollais P, Charnay P,Vyas GN. (1981), "Biology of hepatitis B virus", *Science*, vol. 213,no. 4506, pp. 406-411.

Tiollais P, Pourcel C, Dejean A. (1985), "The hepatitis B virus", *Nature*, vol. 317,no. 6037, pp. 489-495.

Tong MJ, Thursby M, Rakela J, McPeak C, Edwards VM, Mosley JW. (1981), "Studies on the maternal-infant transmission of the viruses which cause acute hepatitis", *Gastroenterology*, vol. 80,no. 5 pt 1, pp. 999-1004.

Tsuchie H, Kurimura O, Tamura I, Shimase K, Kaneto E, Kurimura T, Tsuda F, Mayumi M. (1984), "Incidences of HBeAg and anti-HBe in, and clinical course of hepatitis B virus carriers", *Biken Journal*, vol. 27,no. 4, pp. 169-176.

Tsuda F, Naito S, Takai E, Akahane Y, Furuta S, Miyakawa Y and Mayumi M. (1984) "Low molecular weight (7s) immunoglobulin M antibody against hepatitis B core antigen in the serum for differentiating acute from persistent hepatitis B virus infection". *Gastroenterology* 87, pp. 159-164.

Uchida T, Aye TT, Shikata T, Yano M, Yatsushashi H, Koga M, Mima S. (1994), "Evolution of the hepatitis B virus gene during chronic infection in seven patients", *Journal of Medical Virology*, vol. 43,no. 2, pp. 148-154.

Vall Mayans M, Hall AJ, Inskip HM, Chotard J, Lindsay SW, Coromina E, Mendy M, Alonso PL, Whittle H. (1990), "Risk factors for transmission of hepatitis B virus to Gambian children ", *Lancet*, vol. 336,no. 8723, pp. 1107-1109.

Vall Mayans M, Hall AJ, Inskip HM. (1994), "Do bedbugs transmit hepatitis B? ", *Lancet*, vol. 344,no. 8927, p. 962.

van Ditzhuijsen TJ, Selten GC, van Loon AM, Wolters G, Matthyssen L, Yap SH. (1985), "Detection of hepatitis B virus DNA in serum and relation with the IgM class anti-HBc titers in hepatitis B virus infection", *Journal of Medical Virology*, vol. 15,no. 1, pp. 49-56.

van Hattum J, Schreuder GM, Schalm SW. (1987), "HLA antigens in patients with various courses after hepatitis B virus infection", *Hepatology*, vol. 7,no. 1, pp. 11-14.

Viviani S, Jack A, Hall AJ, Maine N, Mendy M, Montesano R, Whittle HC. (1999), "Hepatitis B vaccination in infancy in The Gambia: protection against carriage at 9 years of age", *Vaccine*, vol. 17,no. 23-24, pp. 2946-2950.

Waters JA, Brown SE, Steward MW, Howard CR, Thomas HC. (1992), " Analysis of the antigenic epitopes of hepatitis B surface antigen involved in the induction of a protective antibody response". *Virus Research*, vol. 22, pp. 1-12.

Weissberg JI, Andres LL, Smith CI, Weick S, Nichols JE, Garcia G, Robinson WS, Merigan TC, Gregory PB. (1984), "Survival in chronic hepatitis B. An analysis of 379 patients", *Annals of Internal Medicine*, vol. 101,no. 5, pp. 613-616.

Whittle HC, Bradley AK, McLauchlan K, Ajdukiewicz AB, Howard CR, Zuckerman AJ, McGregor IA. (1983), "Hepatitis B virus infection in two Gambian villages", *Lancet*, vol. 1,no. 8335, pp. 1203-1206.

Whittle HC, Lamb WH, Ryder RW. (1987), "Trials of intradermal hepatitis B vaccines in Gambian children", *Annals of Tropical Paediatrics*, vol. 7,no. 1, pp. 6-9.

Whittle H, Inskip H, Bradley AK, McLaughlan K, Shenton F, Lamb W, Eccles J, Baker BA, Hall AJ. (1990), "The pattern of childhood hepatitis B infection in two Gambian villages", *Journal of Infectious Diseases*, vol. 161,no. 6, pp. 1112-1115.

Whittle HC, Inskip H, Hall AJ, Mendy M, Downes R, Hoare S. (1991), "Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia", *Lancet*, vol. 337,no. 8744, pp. 747-750.

Whittle HC, Maine N, Pilkington J, Mendy M, Fortuin M, Bunn J, Allison L, Howard C, Hall A. (1995), "Long-term efficacy of continuing hepatitis B vaccination in infancy in two Gambian villages ", *Lancet*, vol. 345,no. 8957, pp. 1089-1092.

Wilson AG, de Vries N, Pociot F, di Giovine FS, van der Putte LB, Duff GW. (1993), "An allelic polymorphism within the human tumor necrosis factor alpha promoter region is strongly associated with HLA A1, B8, and DR3 alleles", *Journal of Experimental Medicine*, vol. 177,no. 2, pp. 557-560.

Yang PM., Chen DS, Lai MY, Su IJ, Huang GT, Lin JT, Sheu JC, Hsu HC, Sung JL. (1987), "Clinicopathologic studies of asymptomatic HBsAg carriers: with special emphasis on carriers older than 40 years", *Hepatogastroenterology*, vol. 34,no. 6, pp. 251-254.

Yang Z, Lauder IJ, Lin HJ. (1995), "Molecular evolution of the hepatitis B virus genome", *Journal of Molecular Evolution*, vol. 41,no. 5, pp. 587-596.

Yu MC, Yuan JM, Ross RK, Govindarajan S. (1997), "Presence of antibodies to the hepatitis B surface antigen is associated with an excess risk for hepatocellular carcinoma among non-Asians in Los Angeles County, California", *Hepatology*, vol. 25,no. 1, pp. 226-228.

Zhang X, Zoulim F, Habersetzer F, Xiong S, Trepo C. (1996), "Analysis of hepatitis B virus genotypes and pre-core region variability during interferon treatment of HBe antigen negative chronic hepatitis B", *Journal of Medical Virology*, vol. 48,no. 1, pp. 8-16.